



<https://theses.gla.ac.uk/>

Theses Digitisation:

<https://www.gla.ac.uk/myglasgow/research/enlighten/theses/digitisation/>

This is a digitised version of the original print thesis.

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>  
[research-enlighten@glasgow.ac.uk](mailto:research-enlighten@glasgow.ac.uk)

EJH/

**Identification and phylogenetic analysis of  
morphologically similar naked amoebae  
using the ssrRNA**

Gary Patrick Sims B.Sc.

A thesis in fulfilment of the requirements for the degree of  
Doctor of Philosophy of the University of Glasgow,  
March 1997.

University Marine Biological Station,  
Millport, Isle of Cumbrae, Scotland  
and the  
Division of Infection and Immunity,  
Institute of Biomedical and Life Sciences,  
University of Glasgow, Glasgow, Scotland.

ProQuest Number: 10391271

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10391271

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code  
Microform Edition © ProQuest LLC.

ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 – 1346

Theris  
10840  
Copy 2





## Abstract

Flat, fan-shaped amoebae are common in many aquatic habitats where they play an important role in microbial food webs. This is particularly true for benthic sediments where they are major consumers of bacteria. However, ecological studies are hindered by difficulties associated with detecting and identifying these small organisms. This project was the first molecular study to examine phylogenetic relationships among these morphologically similar amoebae, and to assess the utility of a molecular marker for identification purposes.

Restriction analyses of the small-subunit ribosomal rRNA (ssrRNA) gene from 15 strains demonstrated that there was a high degree of sequence diversity among fan-shaped amoebae, and this form of analysis may prove to be a useful ecological tool for the unambiguous identification of these organisms.

The primary and secondary structure of the ssrRNA gene from one representative fan-shaped amoeba, *Vannella anglica*, was determined. The ssrRNA was 1962 nucleotides in length with a low G+C content of 37.1%. Several regions were recognised which may be targeted by oligonucleotide probes for rapid identification in mixed cultures. Phylogeny inference programs using sequence data suggest that the fan-shaped amoebae branched from the eukaryotic tree before the divergence of the alveolates, fungi, plants and animals. *Vannella anglica* did not associate with *Acanthamoeba* or *Hartmannella* which indicates that the fan-shaped amoebae form a separate amoeboid lineage, and the subclass Gymnamoebia is polyphyletic.

Phylogenetic analysis of 12 morphologically similar fan-shaped *Vannella* and *Platyamoeba*, using partial ssrRNA sequence analysis, suggested that the cell surface structure may not be a rigid diagnostic feature for the classification of these organisms. Moreover, the molecular and ultrastructural data suggest that one common, small, unidentified fan-shaped amoeba is not related to other *Vannella* and *Platyamoeba* species, and probably represents a new genus.

## Table of contents

<b>Title page</b>	1
<b>Abstract</b>	2
<b>Table of contents</b>	3
<b>List of tables</b>	9
<b>List of figures</b>	10
<b>Abbreviations</b>	12
<b>Acknowledgments</b>	14
 <b>CHAPTER 1: General introduction</b>	 15
Objects of research	24
 <b>CHAPTER 2: Isolation and characterization of amoebae ssrRNA</b>	 
<b>genes</b>	25
Introduction	26
Structure and function of the ribosome	26
Loci of nuclear ribosomal RNA genes	26
Isolation of ribosomal genes	27
Characterization of ssrRNA genes	28
Summary of chapter	29
Materials and methods	31
Amoebae isolates	31
Light microscopical examination of amoebae isolates	31
Culturing techniques	31
Extraction of DNA from naked amoebae	33
Guanidium thiocyanate	34
Proteinase K	35
Boiling sodium dodecyl sulphate (SDS)	35
5% Chelex-100	36
10% Chelex-100/SDS	36
Examination of DNA extractions	36
PCR amplification of ssrRNA genes	37
Southern hybridization	38
Restriction analysis of PCR amplified ssrRNA genes	39

Phylogenetic analysis of riboprinting data . . . . .	40
Results . . . . .	41
Light microscopical examination of amoebae isolates . . . . .	41
VP3 . . . . .	41
Species E . . . . .	42
SIA . . . . .	44
Growth characteristics of amoebae cultures . . . . .	44
Agar cultures . . . . .	45
Liquid cultures . . . . .	46
Culture contamination . . . . .	46
Extraction of genomic DNA from amoebae cultures . . . . .	46
PCR development . . . . .	50
Refinements to template DNA . . . . .	50
Reducing stringency . . . . .	51
Alteration of other parameters . . . . .	52
PCR amplification of <i>Acanthamoeba polyphaga</i> ssrRNA . . . . .	52
Verification of the <i>A. polyphaga</i> ssrRNA . . . . .	53
Analysis of GenEMBL <i>A. polyphaga</i> ssrRNA sequences . . . . .	53
Analysis of <i>A. polyphaga</i> riboprints . . . . .	57
Optimisation of the PCR conditions . . . . .	57
Template DNA . . . . .	57
PCR stringency . . . . .	58
DNA polymerases . . . . .	60
Other PCR reagents . . . . .	62
Inconsistency of PCR amplification . . . . .	62
Final PCR conditions . . . . .	62
Amplification of ssrRNA genes from fan-shaped amoebae . . . . .	63
Confirmation of amoeba ssrRNA amplifications by hybridization . . . . .	66
Specificity of the digoxigenin-labelled ribosomal probe . . . . .	66
Hybridization with <i>Vannella</i> ssrRNA PCR amplifications . . . . .	67
Restriction analysis of amoebae ssrRNA genes . . . . .	70
<i>AluI</i> riboprints . . . . .	70
<i>HinfI</i> riboprints . . . . .	73
Use of riboprinting for identification of field isolates . . . . .	76
Phylogenetic analysis of riboprinting data . . . . .	77
Discussion . . . . .	81

Isolation of amoebal DNA . . . . .	81
PCR development . . . . .	82
Amplification of amoeba <i>ssrRNA</i> genes . . . . .	84
Characterization of amoeba isolates using riboprinting . . . . .	85
Identification of field isolates . . . . .	86
Phylogenetic analysis . . . . .	88

### CHAPTER 3: Cloning, sequencing and secondary structure of the *Vannella*

<i>anglica ssrRNA</i> . . . . .	90
Introduction . . . . .	91
Primary sequence of <i>ssrRNA</i> . . . . .	91
Secondary structure of <i>ssrRNA</i> . . . . .	91
Higher-order interactions . . . . .	94
Structure-function relationship . . . . .	95
Ribosomal RNA sequencing strategies . . . . .	95
Cloning versus direct sequencing . . . . .	96
Cloning PCR-amplified genes . . . . .	96
DNA sequencing . . . . .	98
Summary of chapter . . . . .	100
Materials and methods . . . . .	101
TA-cloning . . . . .	101
Preparation of pDK101 <i>Xcm</i> I 3 kb fragment . . . . .	101
Ligation of linear plasmid DNA and PCR products . . . . .	103
Transformation of plasmids and selection of recombinant colonies . . . . .	103
Analysis of colonies with suspected recombinant plasmids . . . . .	103
Sub-cloning <i>V. anglica ssrRNA</i> gene . . . . .	104
DNA sequencing . . . . .	104
Preparation of template DNA . . . . .	105
Sequencing reactions . . . . .	105
Gel casting, running conditions and autoradiography . . . . .	107
Primary sequence analysis and secondary structure determination . . . . .	107
Results . . . . .	108
TA-cloning using pDK101 . . . . .	108
TA-cloning using pCRII . . . . .	110
Sub-cloning the <i>ssrRNA</i> gene from pCRV01.106 . . . . .	113

Sequencing . . . . .	.116
Primary ssrRNA gene sequence . . . . .	.117
Secondary structure of the ssrRNA . . . . .	.119
Pseudoknots . . . . .	.121
Lone-pair interactions . . . . .	.121
Non-canonical base-pairing . . . . .	.122
Non-canonical extended helices . . . . .	.122
Discussion . . . . .	.123
TA-cloning with pDK101 . . . . .	.123
TA-cloning with pCRII . . . . .	.124
Composition of <i>V. anglica</i> ssrRNA . . . . .	.125
Secondary structure and higher-order interactions . . . . .	.126
Helix E8-1 . . . . .	.126
Helices E23-1 to E23-5 . . . . .	.126
Bilateral bulge of helix 29 . . . . .	.127
Internal loop of helix 43 . . . . .	.128
‘Helix’ 45-1 . . . . .	.128
Pseudoknots . . . . .	.129
Non-canonical extended helices . . . . .	.131
Non-canonical pairings . . . . .	.132
Fidelity of the ssrRNA sequence . . . . .	.132
Development of the ssrRNA structure . . . . .	.134

## **CHAPTER 4: Molecular phylogeny of the amoebae . . . . .**

Introduction . . . . .	.136
Phylogeny of the protists . . . . .	.136
Classification of amoebae . . . . .	.137
Molecular data . . . . .	.141
Ribosomal RNA sequences . . . . .	.142
Molecular phylogeny of prokaryotes . . . . .	.142
Molecular phylogeny of eukaryotes . . . . .	.143
Molecular phylogeny of the crown taxa . . . . .	.146
Phylogenetic analysis of rRNA sequence data . . . . .	.148
Phylogenetic inference . . . . .	.149
Parsimony methods . . . . .	.149
Maximum likelihood . . . . .	.150

Pair-wise distance methods . . . . .	.151
Bootstrap resampling . . . . .	.151
Summary of chapter . . . . .	.152
Materials and methods . . . . .	.153
Sequences . . . . .	.153
Alignments . . . . .	.153
Exclusion of nucleotide positions . . . . .	.156
Preliminary phylogenetic analyses . . . . .	.167
Estimating the phylogenetic position of <i>V. anglica</i> in the eukaryotic tree . . . . .	.167
Results . . . . .	.169
Preliminary phylogenetic analyses . . . . .	.169
Parsimony analysis . . . . .	.171
Maximum likelihood analysis using FASTDNAML . . . . .	.173
Effect of addition and removal of taxa . . . . .	.175
Tree rearrangements . . . . .	.177
Discussion . . . . .	.178
Preliminary phylogenetic analyses . . . . .	.179
Phylogenetic relationships among the crown taxa . . . . .	.179
Phylogenetic position of <i>Phreatamoeba balamuthi</i> . . . . .	.181
Phylogenetic position of <i>Vannella anglica</i> . . . . .	.181
Polyphyly within the phylum Rhizopoda . . . . .	.184

## CHAPTER 5: Identification and phylogenetic analysis of

### **morphologically similar fan-shaped amoebae using partial ssrRNA sequence data . . . . .**

ssrRNA sequence data . . . . .	.186
Introduction . . . . .	.187
Taxonomy of the fan-shaped amoebae . . . . .	.187
Taxonomic characters for identification . . . . .	.191
Protein analysis . . . . .	.197
Analysis of repetitive DNA . . . . .	.198
Specific analysis of amplified rDNA . . . . .	.200
Use of ribosomal RNA probes . . . . .	.202
Summary of chapter . . . . .	.202
Materials and methods . . . . .	.204
Amplification of partial ssrRNA genes . . . . .	.204

Cloning the ssrRNA V7 V8 variable region . . . . .	.204
Sequencing cloned gene fragments . . . . .	.206
Primary sequence comparisons . . . . .	.206
Phylogenetic analysis of V7 V8 DNA sequences . . . . .	.208
Transmission electron microscopy . . . . .	.209
Results . . . . .	.211
Partial ssrRNA gene amplifications . . . . .	.211
Cloning . . . . .	.214
Partial ssrRNA V7 V8 sequencing . . . . .	.217
Phylogenetic analyses on the complete data set . . . . .	.221
Phylogenetic analyses on a restricted data set . . . . .	.224
Rooting the tree . . . . .	.224
Assessment of the usefulness of diagnostic features . . . . .	.227
Identification of unidentified isolates . . . . .	.229
Examination of ultrastructural features . . . . .	.230
VP3 . . . . .	.230
Species E . . . . .	.234
Discussion . . . . .	.236
Length of the V7 V8 ssrRNA region . . . . .	.236
Secondary structure of expansion segment E45-1 . . . . .	.237
Sequence variation . . . . .	.237
Phylogenetic analyses . . . . .	.238
Re-evaluation of cell surface structure . . . . .	.242
Identification of SIA . . . . .	.244
Identification of VP3 . . . . .	.244
Identification of Species E . . . . .	.245
<b>CHAPTER 6: General discussion . . . . .</b>	<b>.249</b>
<b>References . . . . .</b>	<b>.258</b>
<b>Appendix I: Culture media . . . . .</b>	<b>.300</b>

## List of tables

Table 2.1	Collection of amoebae strains . . . . .	32
Table 2.2	Analysis summary of GenEMBL <i>A. polyphaga</i> ssrRNA sequences . . . . .	56
Table 2.3	Optimal annealing temperatures for PCR amplification of ssrRNA genes . . . . .	61
Table 2.4	Representation of the ssrRNA <i>AluI</i> digestion fragments .	72
Table 2.5	Representation of the ssrRNA <i>HinfI</i> digestion fragments .	75
Table 2.6	Binary data set of <i>AluI</i> and <i>HinfI</i> restriction data . .	78
Table 3.1	Transformation of pDK101 TA-cloned PCR-amplified ssrRNA genes into competent DH5 $\alpha$ cells . . . . .	109
Table 3.2	Transformation of pCRII TA-cloned PCR-amplified ssrRNA genes into competent One Shot and Maximum Efficiency DH5 $\alpha$ I <sup>q</sup> cells . . . . .	109
Table 4.1	Classification of the amoebae . . . . .	138
Table 4.2	Alphabetical listing of eukaryotic ssrRNA sequences . .	154
Table 5.1	The classification of the naked amoebae . . . . .	190
Table 5.2	Listing of morphologically similar <i>Vannella</i> , <i>Platyamoeba</i> , <i>Pessonnella</i> and <i>Clydonella</i> species and unidentified isolates	192
Table 5.3	Primers for the PCR amplification of partial ssrRNA gene sequences . . . . .	205
Table 5.4	Listing of isolates used in partial ssrRNA sequence analyses . . . . .	207



## List of figures

Figure 2.1	Light micrographs of the locomotive forms of unidentified isolates . . . . .	43
Figure 2.2	The lytic properties of DNA extraction methods on <i>Vannella anglica</i> . . . . .	48
Figure 2.3	Extraction of nucleic acid from <i>V. anglica</i> . . . . .	49
Figure 2.4	PCR amplification of the <i>A. polyphaga</i> ssrRNA . . . . .	55
Figure 2.5	Restriction analysis of the <i>A. polyphaga</i> ssrRNA gene . . . . .	55
Figure 2.6	Effect of magnesium chloride concentration on PCR amplification . . . . .	59
Figure 2.7	Small-subunit ribosomal RNA PCR amplifications from fan-shaped amoebae . . . . .	65
Figure 2.8	Small-subunit ribosomal RNA probe hybridization specificity . . . . .	69
Figure 2.9	Hybridization of the ssrRNA probe with <i>Vannella</i> amplifications . . . . .	69
Figure 2.10	Restriction digests of ssrRNA PCR amplifications with <i>AluI</i> . . . . .	71
Figure 2.11	Restriction digests of ssrRNA PCR amplifications with <i>HinfI</i> . . . . .	74
Figure 2.12	Most parsimonious unrooted phylogram for fan-shaped amoebae using ssrRNA restriction data . . . . .	80
Figure 3.1	Secondary structure model for eukaryotic ssrRNA . . . . .	93
Figure 3.2	TA-cloning plasmids with cloning site details . . . . .	102
Figure 3.3	Sub-cloning and sequencing strategy for TA-cloned <i>V. anglica</i> ssrRNA gene . . . . .	106
Figure 3.4	Examination of suspected <i>V. anglica</i> ssrRNA recombinant clones . . . . .	112
Figure 3.5	Sub-cloning the <i>V. anglica</i> ssrRNA gene from pCRV01.106 using <i>EcoRI</i> and <i>BamHI</i> . . . . .	115
Figure 3.6	Primary DNA sequence of the <i>V. anglica</i> ssrRNA gene . . . . .	118
Figure 3.7	Predicted secondary structure of <i>V. anglica</i> ssrRNA . . . . .	120
Figure 3.8	Predicted secondary structure of ssrRNA E43-1 helices . . . . .	130
Figure 4.1	Molecular phylogeny of the eukaryotes (Eukaryota) . . . . .	144

Figure 4.2	Selected eukaryotic ssrRNA sequence alignment . . . . .	.157
Figure 4.3	Preliminary placement of <i>V. anglica</i> in the eukaryotic tree	.170
Figure 4.4	Most parsimonious phylogram for eukaryotic taxa using the ssrRNA sequence data . . . . .	.172
Figure 4.5	Maximum likelihood phylogram for eukaryotic taxa using the ssrRNA sequence data . . . . .	.174
Figure 4.6	Parsimony phylograms examining the effect of the inclusion and exclusion of non-crown taxa . . . . .	.176
Figure 5.1	Electron micrographs of the cell surface structure from characteristic <i>Platyamoeba</i> and <i>Vannella</i> species . . . . .	.196
Figure 5.2	Partial ssrRNA gene amplifications from selected amoebae	.213
Figure 5.3	Partial ssrRNA gene amplifications from the V7 V8 variable region . . . . .	.216
Figure 5.4	Alignment of partial ssrRNA V7 V8 sequences . . . . .	.218
Figure 5.5	Unrooted phylograms of naked, fan-shaped amoebae using sequence data from the complete V7 V8 variable region . . . . .	.223
Figure 5.6	Unrooted phylograms of naked, fan-shaped amoebae using a restricted data set from V7 V8 variable region . . . . .	.225
Figure 5.7	Unrooted, majority-consensus most parsimonious phylogram with the exclusion of the Species E taxon . . . . .	.226
Figure 5.8	Rooted, majority-consensus most parsimonious phylograms	.228
Figure 5.9	Electron micrographs of unidentified isolate VP3 . . . . .	.231
Figure 5.10	Electron micrographs of the fine structure of the cell membrane from <i>Vannella</i> species . . . . .	.233
Figure 5.11	Electron micrographs of unidentified isolate Species E . . . . .	.235

## Abbreviations

A	adenine
AS	amoeba saline
ATCC	American Type Culture Collection
bp	base-pair
CCAP	Culture Collection of Algae and Protozoa
C	cytosine
dATP, ddATP	deoxy- and dideoxyadenylate
dCTP, ddCTP	deoxy- and dideoxycytidylate
dGTP, ddGTP	deoxy- and dideoxyguanylate
DIG	digoxigenin
DNA	deoxyribonucleic acid
dNTP, ddNTP	deoxy- and dideoxynucleotide
DTT	dithiothreitol
dTTP, ddTTP	deoxy- and dideoxythymidylate
EM	electron microscopy
G	guanine
EDTA	ethylenediaminetetra-acetic acid
F	forward (primer)
HCl	hydrochloric acid
IPTG	isopropylthio- $\beta$ -galactopyranoside
kbp	kilobase-pairs
L:B	length: breadth ratio
M	adenine or cytosine; molar
MgCl <sub>2</sub>	magnesium chloride
mRNA	messenger RNA
mtDNA	mitochondrial DNA
N	any nucleotide
NA	nutrient agar
NaCl	sodium chloride
NaOH	sodium hydroxide
NaH <sub>2</sub> PO <sub>4</sub>	sodium di-hydrogen orthophosphate
Na <sub>2</sub> HPO <sub>4</sub>	di-sodium hydrogen orthophosphate
NNA	non-nutrient agar

OD	optical density
PCR	polymerase chain reaction
PME	polymerase match enhancer
PPG	protease peptone and glucose liquid culture
R	reverse (primer)
RAPD	randomly amplified polymorphic DNA
rDNA	ribosomal deoxyribonucleic acid
RNA	ribonucleic acid
RFLP	restriction fragment length polymorphism
rRNA	ribosomal ribonucleic acid
SDS	sodium dodecyl sulphate
SSC	sodium chloride, sodium citrate solution
ssrRNA	small-subunit ribosomal RNA
T	thymine
TBE	tris, boric acid, EDTA buffer
TCA	tri-chloroacetic acid solution
TE	tris, EDTA buffer
TEM	transmission electron microscopy
TEMED	tetra-methylethylene-diamine
Tris	tris (hydroxymethyl)-methylamine
tRNA	transfer RNA
U	uracil
V	variable (region)
W	adenine or thymine
w/v	weight: volume ratio
WWW	world wide web
X-Gal	5-bromo-4-chloro-3-indolyl-D-galactosidase
Y	cytosine or thymine
-ln	negative natural logarithm
<sup>35</sup> S	sulphur-35 isotope

## Acknowledgements

I would most like to thank my supervisors Dr Andrew Rogerson and Dr Robert Aitken for their assistance, advice and encouragement throughout the course of my study - I am forever in their debt.

The work was funded by the University of London, and I would like to thank Professor J. Davenport and Professor A. Wardlaw the directors of the Marine Biological Marine Station Millport and the Microbiology Department, University of Glasgow, respectively for the use of their facilities.

At the University of Glasgow, my special thanks to Dr Catherine Sinclair and Janice Gilchrist for their advice and help with hybridization and DNA sequencing, and to Dr Rod Page and Dr Mike Charleston for their assistance with phylogenetic analyses and the kind use of their computers. I would also like to thank Dr T Martin Embley, Dr Johan De Jonckheere and Susan Brown for protocols, assistance and advice with PCR, and Dr Mitchell L. Sogin for providing the plasmid containing the *Vannella* sp. ssrRNA gene. For my work at the Marine Station, I would particularly like to thank Peter Wilson for technical support and advice, and Shona Magill for her patience and resourcefulness in providing computing assistance. For helpful discussion and advice on culture techniques, electron microscopy and taxonomy, I thank Dr Helen Butler, Evelyn Armstrong and Dr Fiona Hannah. Many thanks to all the other students, technical and administrative staff, past and present, who have provided assistance during my term at UMBSM. Special thanks to Sheila Wheadon for assistance in the preparation of this thesis.

Work aside, I would like to thank all the staff and students at the Marine Station, and friends from Millport for making my time there so enjoyable. The friendships I developed, the numerous memorable moments and the vast quantities of fizzy pop we shared will never be forgotten! Finally, I would like to mention my good friend David Searle, who sadly died during the course of our studies. I wish we could have celebrated the completion of our degrees together in style with a "beer-fest" in the West End, a haggis supper and some of his dreadful music.

## **CHAPTER 1**

### **General introduction**

Amoeba(e) is a vernacular term which is used to describe a diverse assemblage of unicellular eukaryotes which use pseudopodia for movement and capturing prey. Traditionally, amoebae were classified in the phylum Rhizopoda (Levine *et al.*, 1980; Page, 1987). This phylum included organisms with lobose (broad), filose or filiform (thin) pseudopodia and those with pseudopodia which form branched or reticulate anastomosing networks. Cells can be either naked with no observable cell coat, or have a cell coat differentiated into distinct scales or a test (shell). Furthermore, other amoebae such as the Schizopyrenida have flagellate stages, and other forms, such as the slime molds, have complicated life-cycles involving the formation of fruiting bodies and spores.

Given the morphological diversity within the group it was clear that not all amoebae are closely related, and that the 'amoeboid form' is not a useful phylogenetic marker as it probably evolved several times throughout the evolution of the eukaryotes. To improve the classification of these organisms, the Rhizopoda was revised resulting in the removal of six of the eight classes (Margulis *et al.*, 1989). The slime molds, amoeboid-flagellates, and amoebae with pseudopodial networks along with some other unusual organisms were all removed from the phylum as they were thought to represent separate eukaryotic lineages. The remaining two classes of Rhizopoda are the Lobosea, which include the majority of naked amoeba with lobose pseudopodia (subclass Gymnamoebia) and all the lobose testate amoebae (subclass Testacealobosia), and the Filosea, which includes the naked and testate amoebae with filose pseudopodia (Schuster, 1989). However, these subclasses are still heterogeneous with considerable morphological variation, and there is little consensus as to which characters most accurately reflect phylogenetic relationships. As a consequence, further taxonomic revisions are inevitable. The relationships between these rhizopods, other amoeboid forms and other eukaryotic organisms is an area of great uncertainty.

The majority of amoebae encountered are naked, lobose amoebae which are often referred to as gymnamoebae (Page, 1987). This term describes all non-sporulating, naked lobose amoebae, and therefore includes all the Gymnamoebia, but also the schizopyrenids which were originally classified within this subclass, but were later excluded on the basis of significant mitotic and ultrastructural differences (Page & Blanton, 1985).

Gymnamoebae are the most common and ubiquitous amoebae being found in all terrestrial and aquatic habitats. Gymnamoebae are most numerous in soils where they are often considered the most important single group of protozoa (Bamforth, 1980; Fenchel, 1987; Foissner, 1987), and significant numbers have also been found in *Sphagnum* bogs, activated sludge tanks and reed beds (Rogerson, 1982; Ramirez *et al.*, 1993; Rivera *et al.*, 1993). In freshwater environments, naked amoebae may be abundant in the plankton of lakes and ponds (Laybourn-Parry *et al.*, 1991; Arndt, 1993), however, elevated numbers are usually associated with sediments or suspended organic floc material (Baldock & Sleight, 1988; Finlay *et al.*, 1988; Laybourn-Parry & Rogerson, 1993). Similarly in marine environments, although significant numbers of naked amoebae have been recorded throughout the water column (Sieburth, 1981; Anderson & Rogerson, 1995), greater abundances of amoebae are associated with surfaces such as the neustonic layer at the water-air interface, suspended flocs of organic matter, marine snow particles, macroalgae and sediments (Davis *et al.*, 1978; Caron *et al.*, 1986; Rogerson, 1991; Rogerson & Laybourn-Parry, 1992a,b; Butler & Rogerson, 1995).

Abundances in soils, sediments and on suspended organic matter suggest that naked amoebae are particularly well adapted to flourish in these nutrient-rich micro-habitats. Their small size, flattened or limax (monopodial and cylindrical) shape and flexibility allow them to penetrate small interstitial spaces, which exclude larger heterotrophic protozoa (Clarholm, 1981; Butler, 1994). In these niches amoebae graze upon resident bacteria which themselves feed on dissolved organic matter primarily secreted by photosynthetic organisms. Naked amoebae can be the most dominant component of the protist community and the primary consumers of bacteria (Vargas & Hattori, 1990; Butler & Rogerson, 1995), playing an important role in the microbial food web by recycling carbon and mineral nutrients (Coleman, 1994; Butler & Rogerson, 1995).

Despite the potential importance in microbial food webs, naked amoebae have been largely 'over-looked' in a majority of ecological studies on benthic and pelagic protists with investigators concentrating on dinoflagellates, flagellates, ciliates and diatoms which are easier to observe (Sieburth, 1979). Using direct counting techniques, these organisms are conspicuous in both fresh and fixed samples, whereas naked amoebae are virtually invisible. Gymnamoebae are small, translucent, slow-moving and usually intimately associated with particulate or



organic matter which renders them very difficult to detect by direct methods (Rogerson, 1991). Moreover, fixatives result in shrinkage or distortion of cell membranes making amoebae unrecognisable (Burnett, 1979). To date, the only effective method for the enumeration of gymnamoebae is enrichment culture, although, this method may significantly underestimate abundances and species diversity (Rogerson & Laybourn-Parry, 1992a).

There have been relatively few surveys on the diversity of gymnamoebae because of a general reluctance of investigators to identify these organisms. Nevertheless, in soil and freshwater habitats the predominant naked amoebae are limax forms such as *Hartmannella* and the schizopyrenids, *Naegleria* and *Vahlkampfia*, and compressed or flattened forms belonging to the *Acanthamoeba*, *Mayorella* and *Thecamoeba* genera (Bamforth, 1980). In pelagic estuarine and full salinity waters the composition of genera change; flattened amoebae with either extended lobose pseudopodia or subpseudopodia such as *Neoparamoeba* and *Mayorella*, and flat, fan-shaped *Platyamoeba* and *Vannella* are most common (Rogerson & Laybourn-Parry, 1992a; Anderson & Rogerson, 1995). Cysts of freshwater *Acanthamoeba* species may also be abundant, although locomotive forms are not found in the marine environment and they do not excyst and grow on marine media (Page, 1983). On the surfaces of macroalgae, 27 morphotypes of naked amoebae have been recognised with *Flabellula*, *Neoparamoeba*, *Stygamoeba* and *Vannella* the most common genera (Rogerson, 1991).

In marine sediments of the Gulf of Mexico and the Atlantic Ocean off the coast of the USA, 26 species from 12 genera were recognised, and several more isolates were recognised to the genus level (Sawyer, 1980). The most frequently encountered naked amoebae were *Paramoeba* (now included in the genus *Neoparamoeba*), fan-shaped *Clydonella* and *Platyamoeba*, *Vexillifera*, *Stygamoeba* and small unidentified limax amoebae. In marine sediments in the coastal waters of the Firth of Clyde, Scotland, a similar composition of naked amoebae was found with *Neoparamoeba*, small fan-shaped *Vannella*- and *Platyamoeba*-like isolates, *Stygamoeba*, and *Hartmannella*, the most commonly isolated (Butler, 1994). Of approximately 60 different morphotypes described in this survey, several did not correspond to any published descriptions of marine species, and seven could not even be tentatively identified to the genus level (Butler, 1994).

Much of the difficulty associated with the identification of gymnamoebae centres around the high number of small, morphologically similar isolates. In marine sediments as many as 61% of the amoebae isolated were smaller than 10  $\mu\text{m}$  and fan-shaped *Vannella*, *Platyamoeba* and *Clydonella*, or limax-like *Hartmannella* (Sawyer, 1980; Butler, 1994). High proportions of naked amoebae of this size were also present in the water column and on the surfaces of seaweeds (Rogerson, 1991; Rogerson & Laybourn-Parry, 1992a; Anderson & Rogerson, 1995). Since these small amoebae are very abundant, have faster growth rates and greater metabolic activities compared to larger amoebae (Butler & Rogerson, 1996), they may well be the primary consumers of bacteria in sediments and on other surfaces in aquatic environments. However, very few species less than 10  $\mu\text{m}$  in size have been described, primarily because there are no rigid diagnostic features on which to base descriptions. Development of our understanding of the role of these amoebae in microbial food webs relies on improving culture techniques and examining the diversity of these organisms and improving methods of identification.

Identification primarily relies upon examination of morphological features using light microscopy. The most important features are the type of movement, whether eruptive or steady flow, the type of pseudopodia or subpseudopodia, the overall shape, and the form of any posterior trailing cytoplasm or uroid during locomotion. The nuclear structure and behaviour of the nucleus during mitosis, observations of the shape and length of pseudopodia of the floating or 'rayed' form, and the occurrence of flagellate stages and cysts may also be useful (Page, 1987). Examination of these features allows keys to be used for the identification to the level of genus for most soil, freshwater and marine naked amoebae (Page, 1983, 1988).

To identify an isolate to species level can be easy or virtually impossible. In many cases there are no rigid diagnostic features on which an accurate identification can be based. Often subtle differences in continuous characters such as size and shape must be used for identification. However, there is often considerable variation within a culture of the same species, so it is therefore necessary to examine the 'average' characteristics of a population (Page, 1983). To ensure a culture is composed of a single species, a pure clonal culture must first be obtained. If the isolate grows on agar this may be accomplished by transferring small agar fragments to fresh media, however, most smaller amoebae only grow in liquid

culture (Butler, 1994). Isolating a single morphotype from a mixed liquid culture can be an arduous task requiring multiple rounds of dilution and subculture. The appearance of cells in a culture may also vary as a function of the culture conditions. To some extent this problem has been remedied by the introduction of a standard set of conditions for observation by light microscopy (Page, 1983), but due to different culture requirements and growth rates, problems still persist. Identification is further complicated by poor illustrations, incomplete descriptions, and the large proportion of small isolates which have yet to be described.

Methods for identifying naked amoebae have progressed with the use of transmission electron microscopy. Electron microscopy has identified significant cytological differences among amoebae with respect to the mitochondrial cristae and the differentiation of the Golgi apparatus (Page & Blanton, 1985). Detailed examination of the cell surface structure, otherwise known as the glycocalyx, may be useful for discriminating between genera which appear very similar by light microscopy. For instance there are insufficient diagnostic features to positively distinguish between the morphologically similar *Mayorella* and *Dactylamoeba*. Both genera have short blunt, non-furcate subpseudopodia usually produced from the anterior hyaloplasm (non-granular cytoplasmic region), but their cell surface structures differ; *Mayorella* have a thick cuticle, whereas *Dactylamoeba* have boat-shaped lattice-work scales (Cann, 1981; Pennick & Goodfellow, 1975). Similarly, the common flat, fan-shaped amoebae without subpseudopodia or a trailing uroid are represented by three genera; *Vannella*, *Platyamoeba* and *Clydonella*. It is particularly difficult to distinguish between marine isolates of these genera, but electron microscopy has shown that *Platyamoeba* have a 'fuzzy' glycocalyx whereas the glycocalyx of *Vannella* is differentiated into distinct glycostyles (Page & Blakey, 1979). However, considering the inherent variation in the structure of the glycocalyx, and the lack of ultrastructural information about the cell coat of *Clydonella* and many *Vannella* and *Platyamoeba* species, there are good grounds to query the validity of this diagnostic feature, and the taxonomic divisions.

Examining evolutionary relationships between taxa can be difficult since there is often a lack of comparable phenotypic characters. Those characters available are usually continuous, and phenotypes may be misinterpreted due to parallelism or convergent evolution (Olsen & Woese, 1993). The relative importance of morphological, physiological and biochemical characters is unclear (Ragan &

Chapman, 1978), and these characters may be affected by environmental factors (Leonardi *et al.*, 1991, 1995). Most problems associated with inferring phylogenies can be overcome by using DNA sequences. Since DNA is the heritable material it may be considered a more important source of phylogenetic information than other character types. There are plenty of individual molecular markers which occur universally allowing phylogenies to be examined from diverse organisms which lack common features. Sequences of DNA are not affected by environmental or physiological factors and the character states are discrete, not continuous. A large number of characters can be examined from a single locus which allows a quantitative analysis of phylogenetic relationships for a whole array of diverse organisms simultaneously (Schlegel, 1991).

With the development of the polymerase chain reaction (PCR) and DNA sequencing protocols the accessibility of molecular data has increased enormously in recent years (Saiki *et al.*, 1988; Embley, 1991; Tracey & Malcahy, 1991). Consequently, molecular techniques are being extensively employed for examining phylogenetic relationships between distantly and closely related organisms, and for identification purposes (e.g. Olsen *et al.*, 1986; Amann *et al.*, 1990; Loomis & Smith, 1990; Hasegawa *et al.*, 1993; Van De Peer *et al.*, 1993). Initially, molecular phylogenies were inferred from protein sequences, such as the cytochrome C gene (Fitch & Margoliash, 1967), but these have been largely superseded by the use of ribosomal RNA gene sequences, particularly the small-subunit ribosomal RNA (ssrRNA) gene (Lane *et al.*, 1985; Schlegel, 1991; Olsen & Woese, 1993). The ssrRNA is chimeric in nature composed of conserved regions of sequence which are important for the structure and function of the rRNA molecule, and variable regions where sequence conservation is less important (Gutell *et al.*, 1992). Primers complementary to the highly conserved regions can be used for PCR amplification and DNA sequencing (Elwood *et al.*, 1985; Medlin *et al.*, 1988). Sequences are aligned in order to ensure that only homologous nucleotide positions are compared and a variety of computer inference programs can be used to produce a phylogenetic tree (Felsenstein, 1993; Swofford, 1993; Swofford *et al.*, 1996). The conserved regions can be used to examine phylogenetic relationships between distantly related taxa, and more variable regions can be used for examining relationships between closely related taxa (Sogin & Gunderson, 1987). Alternatively, restriction analysis of the gene can be used for identification or phylogenetic analysis. In this instance, variation in the restriction sites or the pattern of restriction fragments can be used to detect DNA sequence

differences as a means of examining relationships between closely related taxa or identification of morphologically similar organisms (Clark, 1992; Dowling *et al.*, 1996).

The *ssrRNA* gene has been influential in our understanding of the diversity and phylogenetic divisions among the prokaryotic organisms (Fox *et al.*, 1980; Woese *et al.*, 1990), and the evolution of the eukaryotic cell (Lake, 1988; Sogin, 1991). Comparisons of eukaryotic *ssrRNAs* have shown that eukaryotes have evolved over a long period during which many protist lineages have diverged from the eukaryotic tree before the relatively recent evolution of the plants, animals and fungi (Van De Peer *et al.*, 1993; Schlegel, 1994b). Small-subunit rRNA phylogenies confirmed mitotic and ultrastructural studies which suggested that amoebae were polyphyletic in origin (Page & Blanton, 1985; Clark & Cross, 1988). The schizopyrenids or amoeboid-flagellates branched deep in the eukaryotic tree along with early flagellate groups (Clark & Cross, 1988; Hinkle & Sogin, 1993), whereas *Acanthamoeba* and *Hartmannella* (subclass Gymnamoebia) branch much later with the fungi, plants and animals (Gunderson & Sogin, 1986; Weekers *et al.*, 1994). Furthermore, two naked, anaerobic amoebae *Entamoeba* and *Phreatamoeba* have independent lineages (Hinkle *et al.*, 1994), as do the plasmodial and cellular slime molds (Johansen *et al.*, 1988). More recently, the filose testate amoebae *Paulinella* and *Euglypha* represent another amoeboid lineage (Bhattacharya *et al.*, 1995).

As well as the global phylogeny of the eukaryotes, *ssrRNA* sequence analysis has been used to examine relationships within specific groups. Sequence analyses of microsporidia (Zhu *et al.*, 1994), trichomonads (Gunderson *et al.*, 1995), ciliates (Schlegel *et al.*, 1991), stramenopiles (Leipe *et al.*, 1994), fungi (Van de Peer *et al.*, 1992), plants and green algae (Wilcox *et al.*, 1993), and metazoa (Kobayashi *et al.*, 1993) have shown that each group is monophyletic. To examine the phylogenetic relationships between more closely related taxa requires a sufficient degree of sequence variation. In some instances, *ssrRNA* sequences are too similar to allow phylogenetic inferences to be made (Johnston *et al.*, 1993; Lumb *et al.*, 1993; James *et al.*, 1994), whereas in other taxa sufficient differences can be detected by restriction analysis of PCR-amplified *ssrRNA* genes (riboprinting) or partial sequence analysis (Johnson *et al.*, 1990; Rowan & Powers, 1992; Brown & De Jonckheere, 1994; De Jonckheere, 1994b; Clark *et al.*, 1995).

Restriction analysis has been particularly useful for the rapid identification of morphologically similar species. It has been used to distinguish between non-pathogenic and pathogenic *Entamoeba* (Clark & Diamond, 1991a,b), and for defining species limits within the *Gracilaria*, *Vahlkampfia*, *Trypanosoma* and *Naegleria* genera (Schofield *et al.*, 1991; Brown & De Jonckheere, 1994; Clark *et al.*, 1995; De Jonckheere & Brown, 1995). Alternatively, unique *ssrRNA* sequence elements can be utilised to identify organisms. Primers complementary to unique sequence elements can be used to amplify gene fragments in PCR-based assays (Vodkin *et al.*, 1992), or *rRNA* probes, labelled with fluorochromes or haptens, may be used to identify specific prokaryotic or eukaryotic microorganisms via *in situ* hybridization of whole cells (Amann *et al.*, 1990; Hahn *et al.*, 1992; Lim *et al.*, 1993; Wagner *et al.*, 1994; Lim, 1996).

The *ssrRNA* therefore has multiple uses for identification and phylogenetic analysis. This molecular marker is particularly useful for the analysis of micro-organisms which are difficult to detect, culture or identify. In this project, the use of the small-subunit ribosomal RNA is utilised for the identification and phylogenetic analysis of common, morphologically similar, naked, lobose fan-shaped amoebae, and to examine the phylogenetic position of these amoebae in the eukaryotic tree.

## Objects of research

1. To assess the usefulness of the restriction analysis of the PCR-amplified *ssrRNA* gene sequences (riboprinting) for identification and phylogenetic inference of morphologically similar fan-shaped gymnamoebae (Chapter 2).
2. To determine the primary and secondary structure of the *ssrRNA* of a representative fan-shaped amoeba (Chapter 3).
3. To investigate the phylogenetic position of fan-shaped amoebae in the eukaryotic tree, and to examine potential relationships with other amoebae (Chapter 4).
4. To examine the phylogenetic relationships of morphologically similar fan-shaped amoebae using partial *ssrRNA* sequence analysis, and to assess the validity of morphological and ultrastructural taxonomic characters (Chapter 5).

## **CHAPTER 2**

### **Isolation and characterization of amoebae ssrRNA genes**



## Introduction

### **Structure and function of the ribosome**

Ribosomes are large complexes of RNA and proteins which provide the fundamental machinery for protein synthesis, the translation of messenger RNA into polypeptide chains. Eukaryotic and prokaryotic ribosomes are very similar in structure and function. Each is composed of a small subunit and a large subunit which associate together to form the functional ribosome. The small subunit binds the mRNA and tRNAs, and the large subunit catalyses peptide bond formation.

In eukaryotes, the large subunit is composed of three rRNA species (23-28S, 5.8S and 5S), which are characterised by their sedimentation values, and approximately 49 ribosomal proteins. The small subunit has a single rRNA (16-19S), which is more commonly known as the small-subunit ribosomal RNA (ssrRNA), and approximately 33 proteins. In prokaryotes, the 5.8S rRNA is absent and there are fewer ribosomal proteins. Despite the variation in size of the larger rRNAs, the structures are highly conserved and they play a critical role in the catalytic properties of the ribosome. On the other hand, the large number of proteins are less well conserved and many of them are not essential to ribosome function (Noller, 1984; Raué *et al.*, 1990).

### **Loci of nuclear ribosomal RNA genes**

In eukaryotes, ribosomal RNA genes are found in the nucleus, mitochondria and chloroplasts. The complement of rRNA genes in plastids, however, closely resemble those of bacteria and support the endosymbiotic origin of these organelles. The majority of the genes encoding the nuclear rRNAs are located on one or more chromosomes in tandemly repeated units (rDNA repeat). In most cases, the repeat unit consists of a 16-19S, a 5.8S and a 25-28S rRNA gene which are cotranscribed along with internal and external transcribed spacers, and separated by non-transcribed spacers. The number of copies, and the size of the repeat unit varies from taxa to taxa, but there are usually estimated to be 50-400 chromosomal copies per haploid genome, and the repeat unit is normally 6-13 kbp in length (Mandal, 1984). However, as many as 6,400 copies have been reported in yeast (Appels *et al.*, 1980), and the size of the repeat unit in rat is 44 kbp (Stumph *et al.*, 1979). The 5S rRNA gene is almost always located at other sites in the genome, however, in the slime mold *Dictyostelium* and the yeast *Saccharomyces* this gene is also linked to

the rDNA repeat unit (Cockburn *et al.*, 1978; Bollon, 1982).

In some instances many rDNA repeats are not integrated into chromosomes but exist as extrachromosomal DNA. In the ciliate *Tetrahymena pyriformis* only a single copy of the rDNA repeat is located on a chromosome in the nucleus, but as many as 200 copies may be present on extrachromosomal rDNA (Karrer & Gall, 1976). The extrachromosomal rDNA is in the form of palindromic molecules, each containing two transcription units which are 19 kbp in length, most of which exist as linear molecules, and the remaining ones exist as circles (Engberg, 1985). Similarly, large numbers of extrachromosomal rDNA genes occur in other ciliates such as *Paramecium* and *Stylonychia* (Findlay & Gall, 1978; Lipps & Steinbrück, 1978), as well as the slime molds *Physarum* and *Dictyostelium* (Vogt & Braun, 1976; Cockburn *et al.*, 1978).

During interphase in *Physarum*, the extrachromosomal rDNA are located within the nucleolus, and their replication occurs later than that of the chromosomal DNA (Seebeck & Braun, 1982). In *Amoeba proteus*, the large nucleus has hundreds of peripheral nucleoli which contain a fraction of the rDNA replicating later than the bulk of the nuclear DNA (Minassian & Bell, 1976), which suggests that naked lobose amoebae may also have extrachromosomal rDNA.

### Isolation of ribosomal RNA genes

A variety of strategies have been used to isolate ribosomal RNA gene sequences. In early studies, rRNA genes were isolated from genomic DNA libraries (McCarroll *et al.*, 1983; Gunderson *et al.*, 1987). Genomic DNA libraries were prepared by digesting genomic DNA using restriction enzymes and cloning virtually all the organism's DNA into an appropriate vector, usually a lambda bacteriophage derivative. To identify phage containing rRNA genes, total RNA was isolated, and the rRNA fraction was radioactively labelled and hybridized with the genomic library. Alternatively, a DNA copy of the rRNA gene was synthesized from RNA using reverse transcriptase. However, the preparation of a genomic DNA library was very time consuming, and the use of reverse transcriptase introduced a high frequency of errors into the gene sequence. More recently, the methodology for isolating specific genes has significantly improved with the development of primer-directed enzymatic amplification of DNA sequences using the polymerase chain reaction (Saiki *et al.*, 1985, 1988; Mullis *et al.*, 1986).

The polymerase chain reaction (PCR) is an extremely sensitive technique which allows the amplification of single copy genes from genomic DNA using primers which are complementary to flanking DNA sequence. The technique involves repeated cycles of heat denaturation of the DNA, annealing of the primers to the complementary sequences, and extension from the annealed primers with DNA polymerase. The primers anneal to each strand in the target duplex and are orientated so that DNA extension proceeds across the intervening region. The extension products act as templates for primer annealing in successive cycles of amplification. This cycling results in the exponential accumulation of target sequence and the potential production of millions of copies of target sequence within 30 amplification cycles (Innis *et al.*, 1990).

PCR was first used to amplify the human  $\beta$ -globin gene (Saiki *et al.*, 1985), but has since been used extensively to isolate single copy genes and ribosomal RNA genes. A previous analysis of 45 eukaryotic small-subunit ribosomal RNA (ssrRNA) sequences identified conserved regions at the 5' and 3' termini of the gene which could serve as primers for PCR amplification (Medlin *et al.*, 1988). Primers complementary to these conserved regions, with restriction sites at the 5' ends to facilitate cloning, were used to amplify the ssrRNA gene from the marine diatom *Skeletonema costatum* (Medlin *et al.*, 1988). These primers, and minor modifications of these primers, have since been used to amplify specifically nuclear ssrRNA genes from hundreds of different eukaryotic organisms (Neefs *et al.*, 1993). However, there is no standard set of conditions which will amplify ssrRNA genes from all eukaryotic organisms.

### **Characterization of ssrRNA genes**

The size of eukaryotic nuclear ssrRNA varies from 1.45 kb in the microsporidium *Vairimorpha necatrix*, which has several universal and eukaryotic specific helices missing (Vossbrinck *et al.*, 1987), to greater than 2.4 kb in some insects (Kwon *et al.*, 1991; Chalwatzis *et al.*, 1994), although the size is more typically 1.8 to 1.9 kbp (Neefs, *et al.*, 1993). The size of the ssrRNA is usually similar between species and genera (James *et al.*, 1994; Gunderson *et al.*, 1995), but this trend is not universal. There may be considerable differences among even closely related organisms: expansion segments in hyper-variable regions, or group I introns, can result in large differences in size (Wild & Sommer, 1980; De Jonckheere, 1993; Gast *et al.*, 1994a; Hinkle *et al.*, 1994). These observations, and size comparisons of the gene in

phylogenetic studies, have demonstrated that size can not be used to infer relationships among organisms (Gunderson & Sogin, 1986).

Characterization of the *ssrRNA* genes is thus better founded upon the primary sequence composition. The most comprehensive method is to determine the complete primary structure of the gene by DNA sequencing. Several strategies for sequencing the complete gene have been developed and are described in Chapter 3. Sequencing the entire gene is a laborious and expensive exercise and is impractical if large numbers of genes from similar isolates are to be examined. In these circumstances, partial DNA sequencing of the *ssrRNA* genes has been used to examine phylogenies and to revise controversial taxonomic issues (Johnson *et al.*, 1990; Fenger *et al.*, 1994; De Jonckheere, 1994b). Alternatively, restriction analyses may be used to sample the sequence composition of the *ssrRNA* gene. This may be accomplished by ribotyping, where genomic DNA is digested and the ribosomal gene restriction fragments are detected by hybridizing with a *ssrRNA* probe (Clark *et al.*, 1989). This method, however, has been generally succeeded by riboprinting, where amplified *ssrRNA* genes are digested and the restriction fragments are examined directly by agarose gel electrophoresis (Clark, 1992; Clark & Diamond, 1991a). Riboprinting has been used to differentiate between morphologically similar isolates of the human gut pathogen *Entamoeba histolytica* (Clark and Diamond, 1991ab), and to examine the phylogenies of Anuran trypanosomes (Clarke *et al.*, 1995), the red alga genus *Gracilaria* (Schofield *et al.*, 1991) and schizopyrenids such as *Naegleria* (De Jonckheere, 1994a) and *Vahlkampfia* (Brown & De Jonckheere, 1993). Riboprinting is far quicker and less expensive than DNA sequencing, but still provides useful information for identification and phylogenetic analysis of closely related organisms.

### Summary of chapter

In this chapter, methods are described for the cultivation, DNA extraction and PCR amplification of nuclear *ssrRNA* genes from morphologically similar isolates of naked, flat fan-shaped amoebae. These isolates include 16 isolates from the Culture Collection of Algae and Protozoa (CCAP) assigned to the *Vannella* and *Platyamoeba* genera including several 'authentic' strains, upon which the species diagnoses were based. Three unidentified strains isolated from the field are also described. The isolation of the nuclear *ssrRNA* genes from these strains was confirmed by hybridization with an independent *ssrRNA* probe. The amplified *ssrRNA* gene products from CCAP reference strains were characterized by restriction analysis to

develop a framework for the identification of field isolates, and to provide an indication of the genetic diversity and phylogenetic relationships which may exist between these similar isolates.

## **Materials and methods**

### **Amoebae isolates**

A list of all the morphologically similar flat, fan-shaped amoebae isolates examined in this study, along with details of their date and place of isolation, most detailed reference and the growth media used in their routine cultivation are presented in Table 2.1. All *Vannella* and *Platyamoeba* strains were obtained from the Culture Collection of Algae and Protozoa (CCAP), Institute of Freshwater Ecology, Windermere, Cumbria, England. Strains marked with an asterisk are the 'authentic' strains from which the species diagnoses were first described. Three unidentified flat, fan-shaped isolates VP3, Species E and SIA, which were isolated from field samples by various researchers, were also included in this study. Another naked amoeba *Acanthamoeba polyphaga* was also maintained and utilised in PCR experiments.

### **Light microscopical examination of amoebae isolates**

To confirm the identity of the CCAP cultures, the morphology of the floating and locomotive forms of the amoebae were reexamined in liquid media, using an inverted microscope. The major features such as the length and shape of the floating form pseudopodia, and the general shape, size of the maximum dimension, and the length: breadth ratio of the locomotive form were all scrutinised using recommended methods (Page, 1983). Complete descriptions of microscopical observations were compiled for the three unidentified isolates.

### **Culturing techniques**

Amoebae were maintained on agar or in liquid media. The marine and brackish water isolates were maintained on malt yeast agar containing 75% seawater (MY75S) in petri dishes or in sterile filtered seawater with a sterile rice grain (SW+R) in 50 ml culture flasks. The freshwater cultures were maintained on non-nutrient agar (NNA) containing amoeba saline (AS), except *A. polyphaga* which was maintained axenically in amoeba saline liquid culture containing protease peptone and glucose (PPG). All agar subcultures were supplemented with a thin spread of *Escherichia coli* strain CC118 which was grown on nutrient agar (NA). The recipes for all media are given in Appendix I.

**Table 2.1** Collection of amoebae strains.

Isolate	Number	Isolator	Habitat/Location	Reference	Media
<i>Acanthamoeba polyphaga</i>	1501/3A	Page, 1979	freshwater pond, Madison, Wisconsin, USA.	Page, 1967	PPG
<i>Vannella platypodia</i>	1589/2	Page, 1964	freshwater creek, Indiana, USA.	Page, 1968	NNA
<i>Vannella simplex</i>	1589/3	Huelsman, -	freshwater pool, Botanical Gardens, Bonn, Germany.	Bovee, 1965	NNA
<i>Vannella devonica</i> *	1589/5	Page, 1977	Kinsbridge Estuary, Devon, England.	Page, 1979a	MY75S
<i>Vannella arabica</i> *	1589/7	Page, 1977	seawater tank, Kuwait Inst. Sci. Res., Kuwait.	Page, 1980a	MY75S
<i>Vannella anglica</i> *	1589/8	Page, 1978	brackish water, Walberswick, Suffolk, England.	Page, 1980a	MY75S
<i>Vannella septentrionalis</i> *	1589/10	Page, 1979	River Don Estuary, Scotland.	Page, 1980a	MY75S
<i>Vannella anglica</i>	1589/11	Page, 1978	brackish water, Walberswick, Suffolk, England.	Page, 1980a	MY75S
<i>Vannella lata</i> *	1589/12	Page, 1978	Loch Morar, Scotland.	Page, 1988	NNA
<i>Platyamoeba mainensis</i> *	1565/1	Page, 1969	Damariscotta River Estuary, Maine, USA.	Page, 1971	MY75S
<i>Platyamoeba placida</i> *	1565/2	Page, 1964	freshwater outflow, Madison, USA.	Page, 1968	NNA
<i>Platyamoeba stenopodia</i> *	1656/3	Page, 1967	freshwater lake, Tuskegee, Alabama, USA.	Page, 1969	NNA
<i>Platyamoeba calyotnucleolus</i> *	1656/6	Page, 1972	beach, West Mersea, Essex, England.	Page, 1974a	MY75S
<i>Platyamoeba plurinucleolus</i> *	1565/7	Page, 1972	beach, West Mersea, Essex, England.	Page, 1974a	MY75S
<i>Platyamoeba australis</i> *	1565/9	Page, 1977	Marorehyore, nr. Brisbane, Australia.	Page, 1983	MY75S
<i>Platyamoeba bursella</i>	1565/10	Page, 1977	brackish water, River Deben, Woodridge, Suffolk, England.	Page, 1974a	MY75S
<i>Platyamoeba plurinucleolus</i>	1565/11	Page, 1972	beach, West Mersea, Essex, England.	Page, 1974a	MY75S
VP3	VP3	Rogerson, 1992	off Keppel Pier, Isle of Cumbrae, Scotland.	Chapter 2	MY75S
Species E	VP14	Butler, 1991	marine sediments, Isle of Cumbrae, Scotland.	Butler, 1994 Chapter 2	SW+R
SIA	VP31	Brown, 1994	Lake Windermere, Cumbria, England.	Chapter 2	NNA

Contamination of cultures with other protists was a real hazard. It was imperative that cultures did not become mixed, wrongly labelled or otherwise contaminated with any other eukaryotic organisms. Routine subculturing of agar and liquid culturing was undertaken fastidiously every 3-4 weeks in a Microflow Biological Safety Cabinet (InterMod MDH). Before subculturing, cultures were examined by eye and then by light microscopy to check the culture was homogenous and free from contamination, and to identify regions of dense, healthy amoebae. For agar cultures, an area towards the edge of the petri dish was usually selected as these cells were furthest from the inoculated site. Small agar blocks, approximately 0.5 cm by 0.5 cm, were aseptically transferred to freshly prepared, dry agar plates overlaid with a thin spread of *E. coli*. For liquid cultures, the liquid medium was poured off and the cells attached to the surface of the culture flask were rinsed carefully with fresh sterile liquid media to remove the majority of bacteria and all the rice grains. Amoebae were washed from the flask base with 2 ml of fresh medium and transferred to a fresh culture flask containing approximately 20 ml of liquid medium and a sterile rice grain. All cultures were incubated in the dark at 18-20°C.

### **Extraction of DNA from naked amoebae**

Before the DNA was extracted from amoebae, cultures were examined by light microscopy to ensure that they were not contaminated by any other eukaryotic cell type. In the case of liquid cultures, cells were isolated from dense 3-4 week old cultures by gently pouring off the liquid media, rinsing the flask two times with sterile medium to remove all the rice grains and the majority of the cellular debris and bacteria. Sterile amoeba saline was used for the freshwater isolates and sterile filtered (0.45  $\mu$ m) seawater was used for the marine isolates. The amoebae were then washed into suspension with 2 ml of sterile liquid medium pipetted over the surface of the flask. Amoebae from several liquid cultures were pooled if the cell densities were low. For *A. polyphaga*, the culture flask was simply shaken to suspend the amoebae homogeneously; this species does not adhere to the surface of the culture flask. For agar cultures, cells were harvested from dense 1-3 week old cultures. The agar blocks were removed and the amoeba cells were gently washed into suspension with the appropriate sterile liquid medium. Care was taken not to damage the surface of the agar.



Microscopy was used to ensure that the majority of the amoeba cells had been harvested from liquid and agar cultures. From a dense culture it was possible to isolate approximately  $2 \times 10^4$  amoebae from a single liquid culture and  $1 \times 10^5$  amoebae from an agar culture. Suspensions of amoebae were concentrated by transferring to a 15 ml centrifuge tube and centrifuging at 1000 g for 10 min at 4°C. Supernatants were immediately removed with a pipette taking care not to disturb the cell pellet. The cell pellet was often poorly compacted and a second spin was sometimes necessary.

To find the best DNA extraction method for amoebae, five different methods were compared. For each method the cells were first resuspended with a lytic reagent and incubated to break down the cellular and nuclear membranes. In most instances the nucleic acid was purified with organic solvents, precipitated, rinsed, dried and resuspended in TE buffer (100 mM Tris-HCl pH 7.8, 10 mM EDTA). The five methods, characterized by their major lytic reagent, were guanidine thiocyanate (Pitcher *et al.*, 1989), proteinase K (McLaughlin *et al.*, 1988), boiling sodium dodecyl sulphate (SDS) (Fuhrman *et al.*, 1988), 5% Chelex (Walsh *et al.*, 1991) and 10% Chelex/SDS (De Lamballerie *et al.*, 1992).

### **Guanidium thiocyanate**

Amoeba cells were resuspended in 500 µl of 5 M guanidium thiocyanate, 100 mM EDTA containing 0.5% v/v sarkosyl and incubated at room temperature for 10 min and checked for lysis. The lysate was cooled on ice and 250 µl of ice-cold 7.5 M ammonium acetate was added. The tube was mixed by inversion, left on ice for 10 min and 500 µl of chloroform/pentan-2-ol (24:1) were added after which the phases were mixed thoroughly by inversion. The phases were separated by centrifugation at 14,000 g for 15 min at 4°C. The upper layer was transferred to Eppendorf tubes and 0.54 volumes of ice-cold propan-2-ol were added. Tubes were again mixed by inversion and the DNA was harvested by centrifugation at 10,000 g for 1 min. The DNA pellet was rinsed with 1 ml of ice-cold 70% ethanol, centrifuged at 10,000 g for 8 min at 4°C and the supernatant was removed as described above. The DNA pellet was rinsed a further four times by adding 1 ml of 70% ice-cold ethanol, centrifuging and removing the supernatant. The DNA pellet was dried for 5 min at 50°C and any moisture forming at the neck of the tube was removed by aspiration. The pellet was dried until no further condensation was evident. The pellet was then redissolved in 50 µl of TE buffer and stored at -20°C until required.

### **Proteinase K**

The coll pellet was resuspended in 450  $\mu$ l of 50 mM Tris-HCl pH 7.8, 5 mM EDTA. To the suspension 50  $\mu$ l of 5% SDS and 5  $\mu$ l of 30 mg/ml proteinase K solution were added and the tube was gently mixed and incubated at room temperature for a few minutes. The solution was examined by microscopy to ensure the amoebae had lysed and the lysate was transferred to a 1.5 ml Eppendorf tube and centrifuged at 10,000 g for 10 min.

The supernatant was transferred to a fresh tube and the crude nucleic acid preparation was purified with organic solvents. An equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) was added and mixed by inversion for 1 min, and allowed to stand for 5 min. The phases were separated by centrifugation at 10,000 g for 2 min. The upper aqueous phase was transferred to a fresh Eppendorf tube taking care not to disturb the band of insoluble protein at the aqueous/organic boundary. The volume was made up to 500  $\mu$ l with TE buffer and an equal volume of chloroform was added, mixed by inversion for 1 min, and centrifuged at 10,000 g for 2 min at room temperature. The upper aqueous layer was carefully transferred to a fresh Eppendorf tube.

To precipitate the nucleic acids the tube was cooled on ice and 0.1 volumes of ice-cold 3 M sodium acetate pH 5.2 and 2 volumes of ice-cold absolute ethanol were added. The tube was mixed thoroughly and incubated at  $-80^{\circ}\text{C}$  for at least 30 min or at  $-20^{\circ}\text{C}$  from 2 hours to overnight. The nucleic acids were harvested by centrifugation at 10,000 g for 15 min at  $4^{\circ}\text{C}$ . The supernatant was decanted off, any residual liquid which remained was collected at the bottom of the tube by briefly centrifuging and carefully removed with a pipette. The DNA pellet was rinsed with 1 ml of ice-cold 70% ethanol, centrifuged at 10,000 g for 8 min at  $4^{\circ}\text{C}$  and the supernatant was removed as before. The DNA pellet was rinsed a further time with ice-cold ethanol, centrifuged and the supernatant was removed. The DNA pellet was dried for 5 min at  $50^{\circ}\text{C}$  and any moisture forming at the neck of the tube was removed by aspiration. The pellet was dried until no further condensation was evident, resuspended in 50  $\mu$ l of TE buffer and stored at  $-20^{\circ}\text{C}$  until use.

### **Boiling sodium dodecyl sulphate (SDS)**

Cells were resuspended in 400  $\mu$ l of TE buffer, and 100  $\mu$ l of 5% w/v SDS was added. The suspension was incubated in a boiling water-bath for 5-10 min and

checked for cell lysis. The lysate was centrifuged at 10,000 g for 10 min at room temperature and the supernatant was transferred to a fresh Eppendorf tube. The DNA was purified with organic solvents, precipitated, rinsed, dried and resuspended as previously described for the proteinase K method.

### **5% Chelex-100**

The cells were resuspended in 500  $\mu$ l of sterile distilled water containing 5% w/v Chelex-100 resin beads and incubated in a boiling water-bath for 30 min. The suspension was checked for lysis, vortexed briefly and centrifuged at 10,000 g for 10 min. The supernatant was transferred to a fresh Eppendorf with care not to carry-over any Chelex resin. This crude lysate was stored at -20°C or purified, precipitated, rinsed, dried and resuspended as previously described.

### **10% Chelex-100/SDS**

Cells were resuspended in 500  $\mu$ l of sterile distilled water containing 10% w/v Chelex-100, 0.1% w/v SDS, 1% v/v Nonidet P-40 and 1% v/v Tween 20 and incubated in a boiling water-bath for 30 min. The cells were vortexed briefly and examined for lysis. The lysate was centrifuged at 10,000 g for 10 min, the supernatant was transferred to a fresh Eppendorf tube and this crude lysate was stored at -20°C. Alternatively, the procedure was carried out in the absence of Nonidet P-40 and Tween 20 and the lysate was purified. The DNA was precipitated, rinsed, dried and resuspended as previously described.

### **Examination of DNA extractions**

To assess the quality of the DNA extractions, 5  $\mu$ l of the DNA solution was separated on a 0.7% agarose gel containing 0.5  $\mu$ g/ml ethidium bromide in 0.5x TBE buffer for 1.2 hours at 80 volts and visualized with ultraviolet illumination. The amount and size of the DNA extracted and the relative amount of RNA present was noted. To determine the purity of the extraction, 2  $\mu$ l aliquots of the solutions were diluted to 100  $\mu$ l with distilled water, and the absorbance was determined at 260 and 280 nm using a Ultrospec 450 spectrophotometer (LKB Biochrom) in a 50  $\mu$ l quartz cuvette (Sigma). The average of three consistent optical density readings at these wavelengths was used to determine the  $OD_{260}:OD_{280}$  ratio which produces an arbitrary value for the purity of nucleic acids. Values of 1.8 and 2.0 correspond to pure solutions of DNA and RNA respectively and values significantly less than these suggest that the sample is impure (Sambrook *et al.*, 1989).

### PCR amplification of *ssrRNA* genes

The universal eukaryotic primers used for the amplification of the complete nuclear *ssrRNA* were forward primer 1369 5' AYCTGGTTGATYYTGCCAG 3' and reverse primer 1389 5' TGATCCATCTGCAGGTTACCT 3' (Embley *et al.*, 1992). These oligonucleotides have been modified from the primer set originally used to amplify *ssrRNA* genes (Medlin *et al.*, 1988). Degenerate bases have been introduced to counteract variation in the complementary *ssrRNA* sequence, and the polylinkers have been removed. They are able to amplify almost the entire small-subunit rRNA sequence of most eukaryotes (Neefs *et al.*, 1993). For control *ssrRNA* amplifications it was also appropriate to amplify the *ssrRNA* gene from bacteria using the universal eubacterial primers forward 5' TCAGAWYGAACGCTGGCGG 3' and reverse 5' AAGGAGGTGATCCAGCC 3' (Embley, pers. comm.). All primers were synthesized by either the Biochemistry Department, University of Glasgow, Scotland or Genosys.

Initially PCR amplifications were carried out according to protocols previously described (Embley *et al.* 1992), with minor recommended alterations in the PCR cycling conditions (Embley, pers. comm.). Each PCR amplification was performed in a 500  $\mu$ l Eppendorf tube. The PCR reaction mixture contained 10  $\mu$ l of 10x PCR buffer (670 mM Tris-HCl pH 8.8, 20 mM  $MgCl_2$ ), 25  $\mu$ M of dATP, dCTP, dGTP and dTTP Ultra-pure nucleotides (Pharmacia), 20 pmol of both the forward and reverse universal eukaryotic primers, 0.1 units of Perfect Match Enhancer (Stratagene), 2  $\mu$ l of template DNA solution containing typically 50-150 ng of nucleic acid, and sterile double distilled water to 98  $\mu$ l. The reaction mixture was covered with 100  $\mu$ l of light mineral oil (Sigma). The tubes were preheated to 94°C for 5 min in a MiniCycler Model PTC-150 (MJ Research) and then cooled to 72°C prior to the addition 2  $\mu$ l (1 unit) of *Taq* DNA polymerase (Perkin Elmer). The thermal cycler was programmed for 10 cycles according to the following regime. Denaturation at 94°C for 1 min, annealing at 50°C for 1 min and elongation at 72°C for 2 min. This was followed by 20 cycles of a 30 sec duration at 92°C, 30 sec at 50°C and 72°C for 2.5 min. At the end of the cycling, the reaction was completed with an additional 5 min extension at 72°C. The PCR mixture was transferred from beneath the oil to a fresh tube. Products of each PCR amplification (amplicons) were analysed by electrophoresis on a 0.7% agarose gel.

During the course of the project, the PCR conditions were modified considerably from these initial conditions. Details of the PCR development leading to the optimisation of PCR conditions are given in the Results section of this chapter.

### **Southern hybridization**

To confirm that the amplicons were derived from amoebae *ssrRNA* genes, *Vannella* and appropriate control reaction amplifications were blotted on to nylon membranes and hybridized with a *ssrRNA* digoxigenin (DIG) labelled probe.

The probe was prepared from a pBluescript plasmid containing the complete *Vannella* sp. F-49 ATCC 30945 *ssrRNA* gene, which was kindly donated by Dr M.L. Sogin, Woods Hole, Massachusetts, USA. The *ssrRNA* gene was amplified from the recombinant plasmid, using the universal eukaryotic *ssrRNA* primers, and the 1.95 kbp amplicon was used for the production of a DIG-labelled *ssrRNA* probe. The probe was prepared using the DIG DNA labelling kit (Boehringer Mannheim, Cat.No. 1175 033) exactly as described in the accompanying manual. Briefly, random oligonucleotide hexamer primers were annealed to 300 ng of denatured *Vannella* sp. *ssrRNA* template DNA. The Klenow fragment of DNA polymerase was used to synthesize the complementary DNA strand from the primer with a mixture of deoxyribonucleotides, including DIG-labelled dUTP, to produce DIG-labelled *ssrRNA* gene fragments.

Amplicons from *Vannella* isolates and control reactions derived from both eukaryotic and eubacterial *ssrRNA* primers were separated by agarose gel electrophoresis and blotted by capillary transfer onto positively charged Zeta-probe GT nylon membranes (BioRad) using the alkaline blotting protocol according to the manufacturer's recommendations. The membrane was then separated from the gel, rinsed twice in 2x SSC (0.3 M sodium chloride, 30 mM trisodium citrate, pH 7.0) and air dried. Thereafter, the DNA was fixed to the membrane by drying at 80°C for 40 min on a slab gel drier SGD40 (Appligene).

The *Vannella* and control reaction blots were hybridized with the DIG-labelled *ssrRNA* probe, and hybridization was detected using the DIG luminescent detection kit (Boehringer Mannheim, Cat.No. 1363 514) according to the manufacturer's recommendations. Briefly, the blots were pre-hybridized for 4 hours at 68°C in the absence of probe, hybridized overnight at 68°C with the DIG-labelled *ssrRNA* probe,

and washed in a rotating hybridization oven. To detect hybridization, the blots were initially incubated with anti-DIG Fab alkaline phosphatase conjugate, and then incubated with the chemiluminescent substrate CSPD. The enzyme conjugate bound to the hybridized probe catalysed a luminescent reaction in the presence of CSPD. The luminescence was detected on X-ray film after a 30 min exposure.

### **Restriction analysis of amplified *ssrRNA* genes**

Amplifications producing a single *ssrRNA* gene product were digested with restriction enzymes. If the amplicon was produced in a high yield it was used directly for restriction analysis. Otherwise, the amplicons from several amplifications were pooled, concentrated by precipitating the DNA, and resuspending in a small volume. Twenty microlitres of *ssrRNA* amplicons were digested in a total volume of 40  $\mu$ l with 15 units of *AluI* (Gibco BRL), *HinfI* or *EcoRI* (Sigma) restriction enzyme in the appropriate buffer at 37°C for 2-3 hours. The restriction digests and DNA size standards were loaded onto a 2% agarose gel containing 0.5  $\mu$ g/ml ethidium bromide and the digestion fragments were separated for 4 hours at 80 volts. The riboprints were visualized using ultraviolet illumination and photographed through a red filter.

The size of the digested fragments in the riboprints of *A. polyphaga* CCAP 1501/3A were measured and compared with the expected restriction patterns compiled from information on the complete *A. polyphaga* *ssrRNA* sequences available on the GenEMBL sequence database. A variety of programs from the Genetics Computer Group, Wisconsin Package were used to extract and examine information from the GenEMBL database (Devereux *et al.*, 1984). STRINGSEARCH was used to search the GenEMBL database for *A. polyphaga* *ssrRNA* sequence files. FETCH was used to obtain the sequence information files. FASTA was used to align pairs of *ssrRNA* sequences and to determine their sequence variation. MAP was used to identify restriction sites, and MAPSORT was used to size the expected digestion fragments, so they could be compared to the *A. polyphaga* CCAP 1501/3A riboprints.

For the examination of the *Vannella*, *Platyamoeba* and unknown amoeba riboprints, the size of every restriction fragment for each isolate was estimated using the size standards. The restriction fragments for each isolate were tabulated in columns in descending order of size. Restriction fragments common to different isolates were aligned on the same row of the table so that restriction fragments unique to an

isolate occupy a row to themselves. The similarity or variation in the riboprints for the different isolates, determined by the comparative number of common or unique restriction fragments were analysed.

### **Phylogenetic analysis of restriction data**

For the phylogenetic analysis of the data, a variety of different programs were used from PHYLIP (Phylogeny Inference Package) version 3.5c (Felsenstein, 1993).

Restriction data was first converted into binary data, where the presence (1) or absence (0) of every restriction fragment in descending order of size, was determined for each isolate. The binary data set was resampled using the SEQBOOT bootstrap program to produce 100 data sets. The JUMBLE function was used to randomize the input order of the isolates in each dataset, and parsimony analysis was executed using the Wagner method in the MIX discrete character parsimony program. The consensus most parsimonious unrooted tree was determined using CONSENSE and printed using DRAWTREE.

## Results

### **Light microscopical examination of amoebae isolates**

In general, light microscopical observations of CCAP cultures in liquid media were in good agreement with those from the original species diagnoses, although there was sometimes substantial variation in shape and size, particularly in the case of marine *Vannella* isolates. This variation was overcome by examining populations of amoebae over the entire exponential growth stage (i.e. 2-7 days). Thus 'average' descriptions best resembled the original diagnosis. The full descriptions of the light microscopical observations for the unidentified isolates, and comparisons with similar described species are given below.

### **VP3**

The floating form usually possessed 2-4 hyaline, blunt-ended pseudopodia which were rarely longer than the diameter of the central mass. The floating form did not rapidly anneal to the substratum and take up the characteristic locomotive form. Light micrographs of the locomotive form of VP3 is shown in Figure 2.1 A and B. The shape of the locomotive form was variable, usually flabellate (fan-shaped), with a largest dimension 12-35  $\mu\text{m}$  (mean 22  $\mu\text{m}$ ) and a length: breadth ratio of 0.6-1.2 (mean 0.8). The shape of the posterior edge was variable. The anterior edge was smooth and well rounded. The nucleus was usually clearly visible. The rate of locomotion was  $45 \pm 8.4 \mu\text{m}/\text{min}$ . During locomotion the posterior of the cell was sometimes drawn out and quite narrow. There was no evidence of encystment.

Comparisons of the general size and shape of the locomotive form, and the pseudopodia of the floating form with other marine *Vannella*, *Platyamoeba* and other similarly described isolates suggest VP3 superficially resembles several isolates such as *V. devonica*, *V. anglica*, *V. arabica*, *V. septrentionalis*, *V. sensilis* and *P. bursella* (Page, 1974a, 1979a, 1980a; Bovee & Sawyer, 1979). However, *P. bursella* has transient surface wrinkles, *V. devonica* and *V. septrentionalis* have considerably longer floating form pseudopodia, the length: breadth ratio of *V. anglica* is a little high, and the largest dimension of *V. sensilis* is too small. Therefore, of all the described flat, fan-shaped species, VP3 appears to resemble *V. arabica* the most closely.



### Species E

The floating form was generally rounded with short irregular, blunt pseudopodia. Light micrographs of the locomotive form of Species E are shown in Figure 2.1 C and D. The shape of the locomotive form was variable, usually flabellate or tongue shaped (linguiform). The largest dimension was 6-13  $\mu\text{m}$  (mean 10  $\mu\text{m}$ ) and the length: breadth ratio of 1.0-2.2 (mean 1.5). The flattened hyaloplasm at the anterior, occupied at least two thirds of the cell. The anterior edge was usually well rounded and smooth, although transient folds were sometimes apparent. The dense granuloplasmic mass at the posterior of the cell was either triangular or oval in shape. Within the granuloplasm, the nucleus and vacuoles could not be clearly identified. The locomotive form annealed firmly to the substratum. The rate of locomotion was  $22 \pm 6 \mu\text{m}/\text{min}$ . There was no evidence of encystment.

Relatively few fan-shaped amoebae of this size have been previously described. Sawyer described several *Platyamoeba* isolates from marine surface waters of Chincoteague Bay, Virginia (Sawyer, 1975b), which include small amoebae such as *P. douvresi*, *P. langae*, *P. murchelanoi* and *P. weinsteini*. However, only a single *Vannella* isolate, *V. aberdonica*, has been described which is in this size range (Page, 1980a). Species E can be easily distinguished from all of these aforementioned isolates as the length: breadth ratio of the locomotive form is considerably higher in Species E due to its tendency to form an extended tongue or linguiform shape. Two *Platyamoeba* isolates, the marine isolate *P. leei* and the freshwater isolate *P. stenopodia*, are described as linguiform (Page, 1969; Sawyer, 1975b), although both these isolates are considerably larger than Species E. It was previously suggested that Species E may be a strain of *V. aberdonica* (Butler, 1994).

---

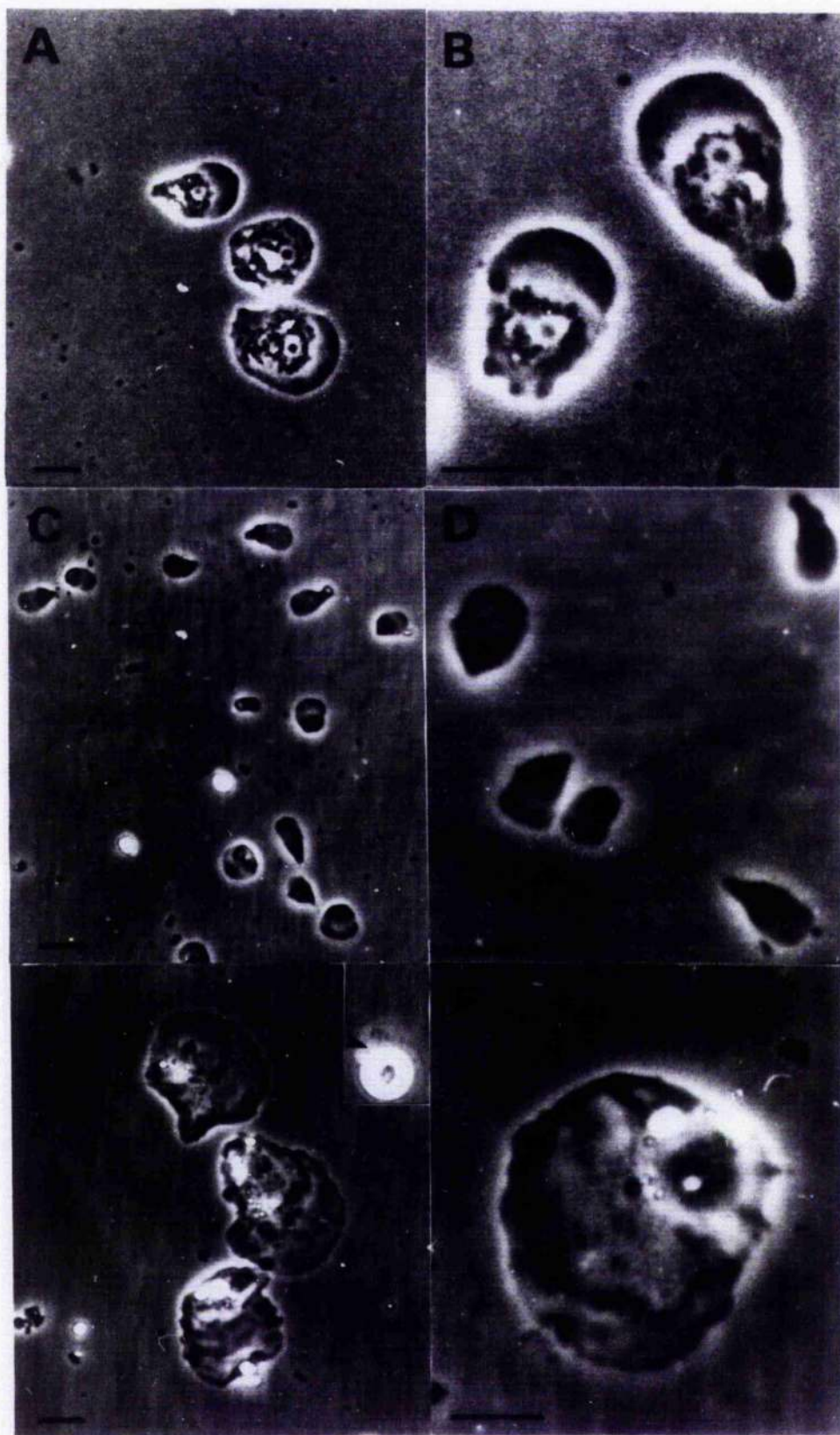
**Figure 2.1** Light micrographs of the locomotive forms of unidentified isolates.

(A,B) VP3

(C,D) Species E

(E,F) SIA with a cyst (c).

Photographs were taken with an inverted Olympus microscope using a phase contrast x40 objective. Scale bar = 10  $\mu\text{m}$  in all cases.



However, unlike *V. aberdonica*, *P. langae* and *P. murchelanoi*, which have no pseudopodia in the floating form, Species E has short irregular pseudopodia. Moreover, the floating form pseudopodia of *P. douvresi* and *P. weinsteini* are much larger than those of Species E. These observations suggest that Species E has not been previously described, but similarities indicate it is likely to be a new species of *Platyamoeba*.

## SIA

The pseudopodia of the floating form were variable, with some pseudopodia short and broad, whereas others were longer than the cell mass and tapering slightly to a blunt end. Light micrographs of the locomotive form of SIA are shown in Figure 2.1 E and F. The shape of the locomotive form was also variable, usually oval sometimes extended, occasionally flabellate. The greatest dimension ranged from 17-35  $\mu\text{m}$  (mean 25  $\mu\text{m}$ ). The length: breadth ratio ranged from 0.6-1.8 (mean 1.1). There were often folds or wrinkles across the hyaloplasm. The anterior edge was sometimes smooth, but often irregular. The granuloplasm was not as granular as many other isolates and the shape was less rigid. The rate of locomotion was  $14 \pm 4$   $\mu\text{m}/\text{min}$ . Cysts, when present, were smooth and approximately 10  $\mu\text{m}$  in diameter (Figure 2.1 E).

SIA is quite distinct from described freshwater *Vannella* species which have floating forms with long, tapering pseudopodia and no cysts. SIA is also quite distinct from *P. stenopodia* which, although of a similar size, is a characteristic extended tongue or linguiform shape (Page, 1969). SIA is also considerably different from *P. schaefferi* which has much larger cysts (Singh & Hanumaiah, 1979), and *Pessonella marginata* which has no cysts (Pussard, 1973). One freshwater '*Vannella*', *Vannella cutleri*, which does have cysts resembles SIA, but this isolate does not have pseudopodia in the floating form (Singh & Hanumaiah, 1979). The only described species that SIA resembles is *Platyamoeba placida* (Page, 1968), which is very similar with respect to the locomotive and floating forms, and cyst morphology.

## Growth characteristics of amoebae cultures

Within this group of amoebae, the growth characteristics were quite variable with respect to growth rate, density and nutritional requirements. However, it was unlikely that detailed descriptions of the growth characteristics of each isolate would be a beneficial diagnostic feature. To aid the maintenance of these cultures

for the production of large numbers of cells for DNA extraction, observations of the culture characteristics of different isolates were useful and some pertinent observations are given below.

### Agar cultures

Amoebae migrated out from the transferred agar block at substantially different rates. Some isolates, such as *V. devonica* and *P. calycinucleolus*, migrated rapidly but divided infrequently resulting in low density cultures covering the entire plate. Other isolates, including *V. arabica* and *P. plurinucleolus*, migrated and divided relatively quickly producing cultures with actively dividing amoebae extending beyond a dense cluster of inactive cells. The freshwater *P. placida* and SIA isolates also migrated and divided quickly, but they encysted when cultures became dense. After four weeks, trophic cells from these cultures were only found around the periphery of the agar plate. Since many cysts were non-viable or may not receive the stimulus to excyst and grow, it was important to initiate subcultures from those peripheral trophic cells. The larger *V. simplex* and *V. lata* isolates tend to have a more methodical migration producing cultures of active cells slowly extending out from the point of transfer. Subculturing and isolating large numbers of cells from these isolates was relatively easy.

A feature of some agar cultures were sites of particularly high cell densities containing tens to one or two hundred cells, against an otherwise low background cell density. At these sites, division may have been rapid, as up to four cells, moving in opposing directions, were still attached to one another as cytokinesis had not been completed. However, the relative size and density of the cells did not indicate that these were sites of rapid proliferation: these dense regions may be a result of aggregation. Nevertheless, these sites of high cell density were useful for subculturing and DNA extraction as they consisted of high numbers of healthy cells which could be easily identified by eye.

With cultures of *P. stenopodia* a further culture characteristic was important when isolating cells for DNA extraction. After a short period in culture this isolate disrupted the integrity of the agar surface either digesting or borrowing into the top few millimetres of the agar. Hence, cells of *P. stenopodia* were not easily washed off and isolated.

## **Liquid cultures**

Small field isolates, like Species E, did not adapt well to growth on agar media, but were well suited to liquid culture where they formed dense cultures and survived for several months by simply replacing the liquid media periodically. This was less time consuming than preparing agar, examining the culture for suitable areas of growth and transferring blocks of agar. Larger isolates, however, do not grow densely in liquid media and obtaining large numbers of amoebae for DNA extraction may require many culture flasks. The use of liquid culture also allowed excess bacteria to be removed from amoebae prior to DNA extraction by repeatedly rinsing the culture flask with fresh liquid media. However, the use of too many rice grains to encourage bacterial numbers facilitated the production of a thin carbohydrate film which coats the culture flask. Some amoebae adhered very tightly to this surface coat and their removal was sometimes difficult.

## **Culture contamination**

Cultures were regularly checked to ensure that they had not been contaminated with another isolate. On occasion in the summer months, cultures did become contaminated with filamentous fungi. Fortunately, however, contamination of this sort was easily identified and any cultures affected were destroyed. No subculturing or DNA extraction was ever undertaken with a culture which was contaminated or suspected of being contaminated.

## **Extraction of genomic DNA from amoebae cultures**

Since the production of large numbers of amoebae for extraction of DNA was not a straight forward task for many of the isolates, several DNA extraction methods were compared to assess which yielded the greatest quantity of DNA from a typical culture of cells.

Several cultures of *V. anglica* were pooled and equal volumes were used to extract DNA by five methods; guanidine thiocyanate, proteinase K, SDS, 5% Chelex-100 and 10% Chelex-100/SDS. The cells were concentrated by centrifugation, and equivalent volumes of cells were examined by light microscopy before and after exposure to the lytic conditions for each DNA extraction technique (Figure 2.2). Freshly harvested cells were numerous before exposure to any lytic extraction reagent (Figure 2.2 A and B), but following incubation with extraction solutions containing guanidine thiocyanate or proteinase K all the cells lysed (Figure 2.2 C and D). With the SDS

treatment, a minority of cells was still visible, although the cell membranes of many of these were disrupted and the nuclear material may have been released (Figure 2.2 E). The lytic properties of 5% Chelex, however, were less potent and the cell membranes remained intact for as many as 10-15% of the cells exposed (Figure 2.2 F). The lytic properties of 10% Chelex-100/SDS were similar to those of SDS.

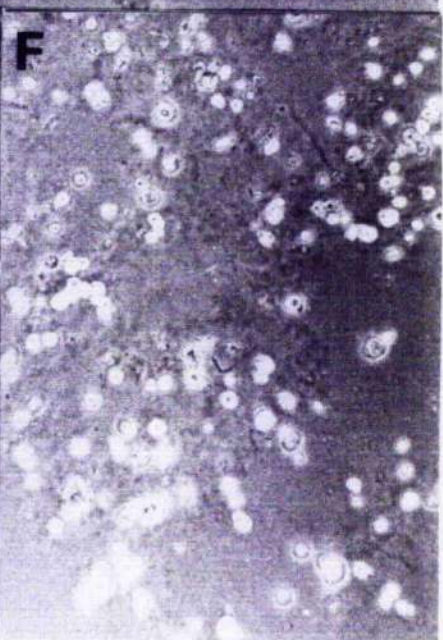
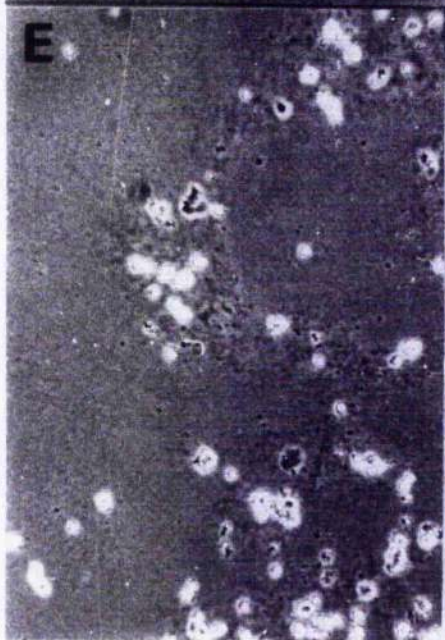
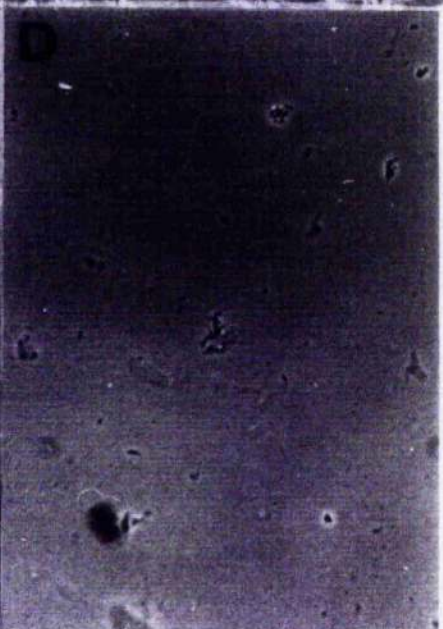
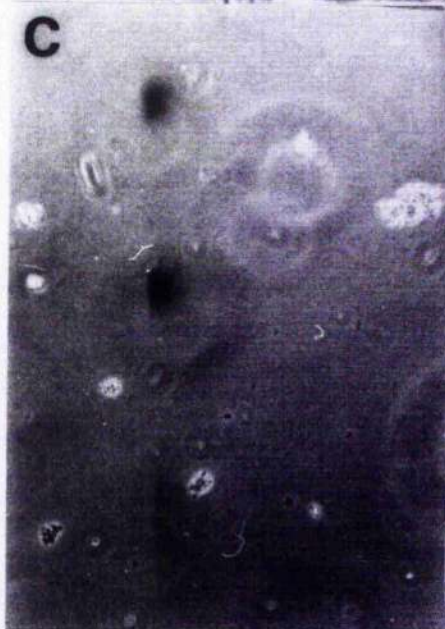
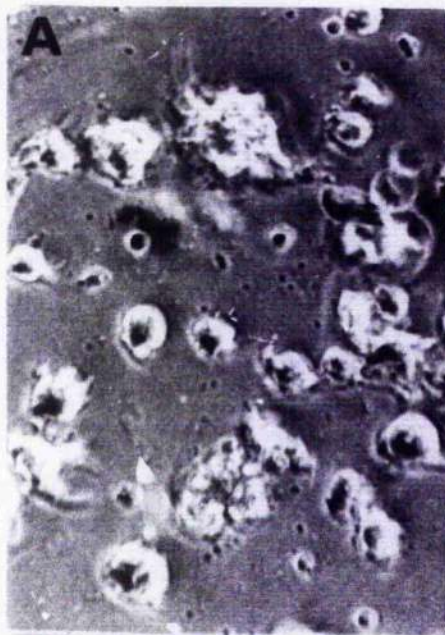
Following cell lysis, the DNA was precipitated and resuspended in TE solution. Aliquots of each DNA extraction solution were separated on a 0.7% agarose gel and visualized under ultraviolet light (Figure 2.3). Different extraction methods yielded markedly different amounts of DNA. The proteinase K and SDS methods produced the most nucleic acid (Lanes 4 and 6) with a bright band at a high molecular weight (>20 kbp) and a faint smear between approximately 0.3 and 2.0 kbp. Treatment with ribonuclease did not affect the high molecular weight band, but it did remove the lower molecular weight smear (Lanes 5 and 7), demonstrating that the high molecular weight band was DNA and the low molecular weight smear was RNA. The remaining methods were far less efficient at extracting nucleic acid. The 10% Chelex-100/SDS method produced a faint DNA band and a faint RNA smear (Lanes 10 and 11), and for 5% Chelex-100 no DNA and only a faint RNA smear (Lanes 8 and 9). Following lysis, the crude extracts were centrifuged so the DNA solution could be removed from the Chelex beads. A substantial volume of solution was

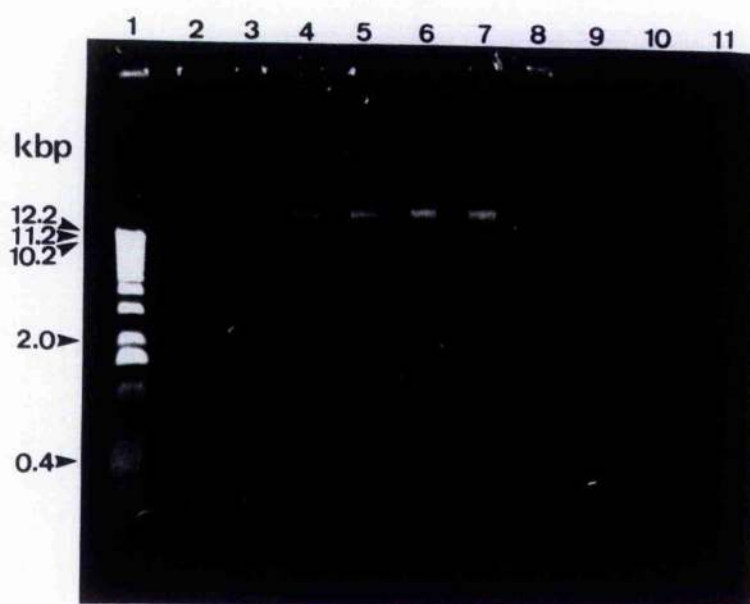
---

**Figure 2.2** The lytic properties of DNA extraction methods on *Vannella anglica*. Equivalent volumes of suspensions of *V. anglica* were examined by light microscopy and photographed before exposure to lytic extraction reagents at 400x and 100x magnification, and after incubation with lytic reagents at 100x magnification.

- (A) before lysis (x400)
- (B) before lysis (x100)
- (C) guanidine thiocyanate
- (D) proteinase K
- (E) SDS
- (F) 5% Chelex







**Figure 2.3** Extraction of nucleic acid from *Vannella anglica*. Nucleic acid was extracted from *V. anglica* using a variety of DNA extraction methods with and without ribonuclease treatment.

Lane 1: 1 kbp ladder

Lane 2: guanidine thiocyanate

Lane 3: guanidine thiocyanate with ribonuclease treatment

Lane 4: proteinase K

Lane 5: proteinase K with ribonuclease treatment

Lane 6: SDS

Lane 7: SDS with ribonuclease treatment

Lane 8: 5% Chelex

Lane 9: 5% Chelex with ribonuclease treatment

Lane 10: 10% Chelex/SDS

Lane 11: 10% Chelex/SDS with ribonuclease treatment



unrecoverable at this stage, but not a sufficient quantity to account for the relatively low quantities of nucleic acid recovered. This suggests that the nucleic acids may have become bound to the Chelex beads. No nucleic acids were recovered from the guanidium thiocyanate method despite the strong lytic properties (Lanes 2 and 3). Precipitation of the nucleic acids for the guanidine thiocyanate method was clearly ineffective. Considering the good yields of DNA prepared using proteinase K and SDS, these methods were chiefly used to extract DNA from amoebae, and no further attempt was made to improve the yields of nucleic acid obtained by the other methods.

The quantity of nucleic acid extracted varied between cultures, although the estimated size of the DNA remained at approximately 20 kbp. In general, most nucleic acid was obtained from freshwater cultures which required a substantial bacterial supplement for dense growth which indicates that a proportion of the nucleic acid extracted was bacterial in origin. Remaining variation in the quantity of nucleic acid extracted can be attributed to the growth rates and densities of cultures. These results demonstrate the importance of removing as many bacteria as possible prior to DNA extraction, to ensure that most of the DNA extracted is amoeba DNA, and the need for multiple cultures of amoebae with low densities.

### **PCR development**

Initial attempts to amplify the *ssrRNA* used conditions characteristic for the amplification of other eukaryotic *ssrRNA* genes with universal primers. These experiments were unsuccessful, but positive amplification of a bacterial toxin gene from a plasmid template consistently produced a single amplification product of the expected 550 bp size in a good yield. Since the target sequence of the positive amplification was quite small and the template DNA was rich in target sequence, this control reaction was likely to be extremely robust. It appeared that amplification of the much larger ribosomal gene from genomic DNA would demand a more specific set of conditions. Therefore, refinements to the PCR conditions were made with respect to the template DNA, stringency of the annealing reaction, PCR reagents and PCR cycling conditions.

### **Refinements to template DNA**

Initial experiments attempted to amplify amoebae *ssrRNA* from serial dilutions of genomic DNA extracted using the proteinase K and SDS methods, as these produced

the greatest yields of DNA. Both these extraction methods include SDS which is a known inhibitor of *Taq* DNA polymerase (Innis *et al.*, 1988). Although DNA purification and rinsing steps were incorporated, it is possible that some SDS may have co-precipitated with the DNA and inhibited the activity of the DNA polymerase. To overcome the possible inhibitory effects, the non-ionic detergents Nonidet P-40 and Tween 20, which can counteract the effects of SDS (Gelfand & White, 1990), were introduced into the PCR mixture, but they failed to improve amplification. The 5% Chelex-100 method, which does not include SDS and is less likely to have inhibitors of PCR (Walsh *et al.*, 1991), was also used directly and following purification, but amplification of the *ssrRNA* remained unsuccessful.

The presence of high levels of RNA in the DNA extraction can also suppress PCR amplification (Pikaart & Villeponteau, 1993), but pretreatment with ribonuclease was also unable to stimulate amplification. The presence of large amounts of bacterial DNA may also have effected PCR, so DNA was extracted from an axenic culture of the naked amoeba *A. polyphaga* CCAP 1501/3A, but the *ssrRNA* gene could not be amplified. A DNA preparation from a species of the amoeba-flagellate genus *Naegleria* was obtained from Dr S. Brown, CCAP, Institute of Freshwater Ecology, Lake Windermere, Cumbria, England. The *ssrRNA* gene was previously amplified from this preparation using identical primers and a similar set of PCR conditions (Brown, pers. comm.). However, the conditions were not appropriate for amplification from this preparation on the thermal cycler used in this study.

### **Reducing stringency**

The stringency of the PCR was reduced in three ways to encourage the annealing of the oligonucleotide primers to the template DNA. Firstly, Perfect Match Enhancer was eliminated from the PCR mixture. Secondly, with a constant final concentration of 2.0 mM magnesium chloride the annealing temperature was reduced stepwise from 50°C through 49°C, 46°C, 43°C and 40°C to 37°C. Finally, maintaining an annealing temperature of 49°C, the final concentration of magnesium chloride was increased from 2.0 mM to 5.0 mM, in 0.5 mM increments. Under the least stringent conditions some amplification products were present, but they were not consistent and they were much smaller than the expected size of eukaryotic *ssrRNA* genes.

### **Alteration of other parameters**

Reducing the reaction volume to 50  $\mu$ l, increasing the primer concentration by up to five-fold and increasing the *Taq* DNA polymerase to 2.5 units per reaction failed to yield products. The denaturing, annealing and elongation times were also increased to 2.5 min, 4.5 min and 7.5 min respectively, but the *ssrRNA* gene could not be amplified from amoeba DNA.

Finally, the default ramping rate of the thermal cycler was reduced during the rise from the annealing temperature to the elongation temperature from the default rate to a 1°C rise every 5 seconds. It was hoped that this would reduce the dissociation of the primers from the template DNA as the temperature increased. A single 2.25 kbp amplicon was produced from *A. polyphaga* template DNA which was characteristic of the size of *Acanthamoeba ssrRNA*.

For this first successful amplification long denaturing, annealing and elongation times were used with a low stringency annealing temperature of just 37°C. Subsequent PCRs demonstrated that it was the high default ramping rate of the thermal cycling machine alone which was responsible for preventing amplification of amoeba *ssrRNA*. Optimisation of the PCR conditions for the amplification of amoeba *ssrRNA* genes and the final PCR conditions are described later in this results section.

### **PCR amplification of *Acanthamoeba polyphaga ssrRNA***

Refinements to the PCR protocol permitted the positive amplification of DNA extracted from *A. polyphaga*. A single product of approximately 2.25 kbp was amplified from DNA extracted using both the proteinase K and the SDS methods (Figure 2.4). The largest *ssrRNA* amplification yields were produced with 10 ng of template for both the proteinase K and SDS methods (Lanes 4 and 8). With 1 ng of template DNA the amplicon yield was reasonable (Lanes 5 and 9), but using 100 pg and 10 pg of template, amplifications were barely detectable. Since the yields for every template dilution were comparable between methods, this shows that the amplification efficiency of amoebae *ssrRNA* genes was independent of the method of DNA extraction, and at least 10 ng of amoebal DNA is required for a good yield of product.

### **Verification of the *A. polyphaga* *ssrRNA***

To verify that the *A. polyphaga* CCAP 1501/3A PCR product was amplified from the nuclear *ssrRNA* gene, the size of the product and *EcoRI* and *AluI* digestion products were compared to the other *A. polyphaga* *ssrRNA* sequences obtained from the GenEMBL database. A STRINGSEARCH of the database identified three *A. polyphaga* *ssrRNA* gene sequences with the accession numbers U07402, U07407 and U07415. A MAP analysis of these sequences was used to locate the *EcoRI* and *AluI* restriction sites. To identify the homologous digestion fragments, a MAPSORT analysis was carried out to obtain the expected sizes of the digestion fragments. FASTA analyses were also performed to determine percentage variation of the *ssrRNA* primary sequence between the three strains. A summary of the MAP and MAPSORT results are shown in Table 2.2, and the *EcoRI* and *AluI* riboprints are shown in Figure 2.5.

### **Analysis of GenEMBL *A. polyphaga* *ssrRNA* sequences**

The length of the three GenEMBL *A. polyphaga* *ssrRNA* sequences were 2,273, 2,276 and 2,287 bp (Table 2.2). Pair-wise FASTA analyses between the three sequences demonstrated that the percentage variability between strains CDC:0884:V029 and JAC/S2 ATCC 50372 was 0.3%, and the percentage variability between both these strains and strain BCM:0173:16 ATCC 50371 was 2.3%. The variation in the size and primary sequence was due to small insertions/deletions and sequence differences spread throughout the gene.

MAP and MAPSORT analyses demonstrated that a single *EcoRI* restriction site was present, and the size of the two restriction fragments was similar for all three strains (Table 2.2). For the *AluI* enzyme, there were seven homologous *AluI* restriction sites, but an additional site was present in strains CDC:0884:V029 and JAC/S2 ATCC 50372 which was absent from strain BCM:0173:16. As a result, the large 682 bp *AluI* fragment in strain BCM:0173:16 was replaced by two smaller fragments of 489 and 185/186 in the other two strains (Table 2.2). This *AluI* restriction polymorphism was detected despite only a 2.3% variation in the primary sequences between strain *A. polyphaga* BCM:0173:16, and the other two strains.

**Figure 2.4** Amplification of the *A. polyphaga* ssrRNA gene.

Lane 1:  $\lambda$  HindIII

Lane 2: positive PCR control.

Lane 3: negative PCR control.

(Lanes 4-7) *A. polyphaga* ssrRNA amplifications from Proteinase K extractions with

Lane 4: 10 ng of DNA

Lane 5: 1 ng of DNA

Lane 6: 100 pg of DNA

Lane 7: 10 pg of DNA

(Lanes 8-11) *A. polyphaga* ssrRNA amplifications from boiling SDS extractions with

Lane 8: 10 ng of DNA

Lane 9: 1 ng of DNA

Lane 10: 100 pg of DNA

Lane 11: 10 pg of DNA

**Figure 2.5** Restriction analysis of the amplified *A. polyphaga* ssrRNA gene.

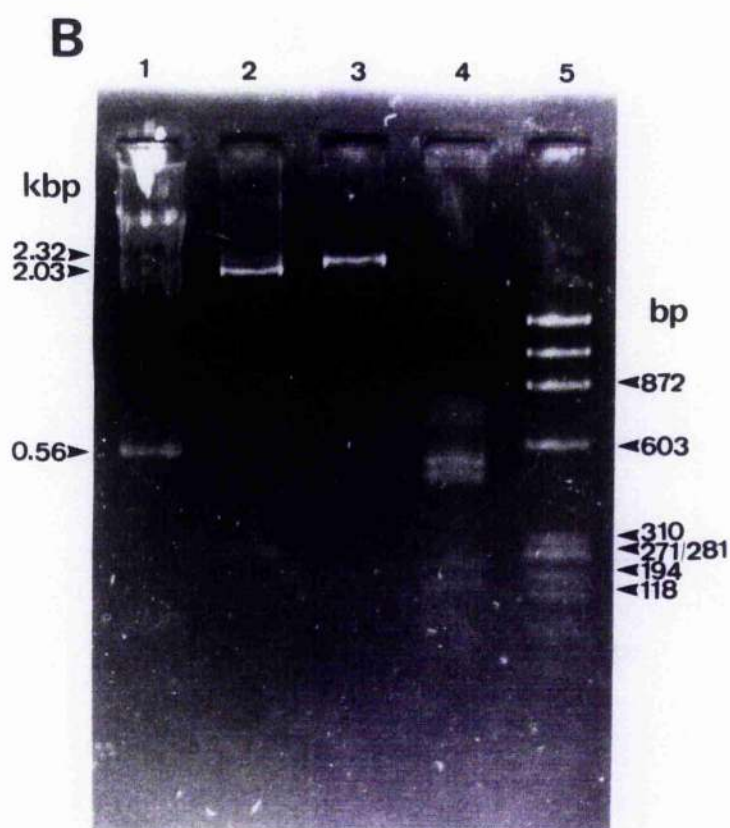
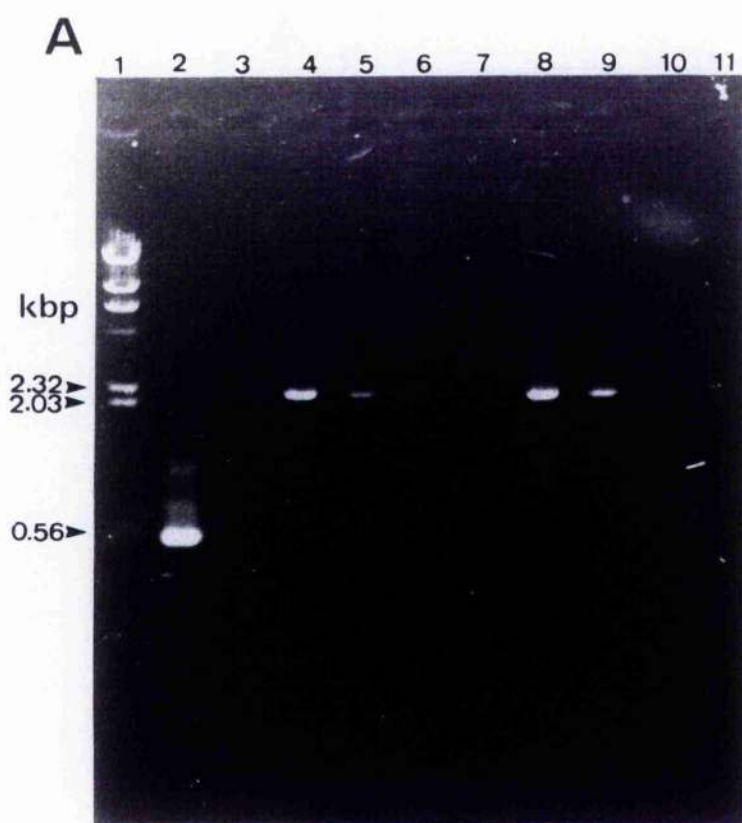
Lane 1:  $\lambda$  HindIII

Lane 2: *A. polyphaga* ssrRNA EcoRI

Lane 3: *A. polyphaga* ssrRNA undigested

Lane 4: *A. polyphaga* ssrRNA AluI

Lane 5:  $\phi$ X174 RF DNA HaeIII



**Table 2.2** Analysis summary of GenEMBL *A. polyphaga* ssrRNA sequences. The sizes of the ssrRNA sequence and the digested fragments are all given in base-pairs. Restriction fragments which are not comparable for all three strains are underlined.

Strain	GenEMBL Accession number	Size of ssrRNA sequence	Size of <i>Eco</i> RI frags	Size of <i>Alu</i> I frags
<i>Acanthamoeba polyphaga</i> CDC:0884:V029	U07402	2,273	2,030 243	522 <u>489</u> 476 217 <u>185</u> 169 108 69 38
<i>Acanthamoeba polyphaga</i> JAC/S2 ATCC 50372	U07415	2,276	2,032 244	523 <u>489</u> 476 218 <u>186</u> 169 108 69 38
<i>Acanthamoeba polyphaga</i> BCM:0173:16 ATCC 50371	U07407	2,287	2,046 241	<u>682</u> 524 480 215 169 110 69 38

### **Analysis of *A. polyphaga* riboprints**

The *A. polyphaga* *Eco*RI and *Alu*I riboprints are shown in Figure 2.5. Digestion of the *A. polyphaga* amplicon with *Eco*RI produced two digestion fragments; a bright band estimated at 2.0 kbp and a weak band estimated at 250 bp (Lane 2). Digestion with *Alu*I resulted in a bands at approximately 680 bp, 530 bp, 490 bp, and weaker bands at 220 bp and 170 bp (Lane 4). The combined size of these fragments was significantly less than the total size of amplicon which suggests that additional smaller fragments were present, although not visible.

The estimated restriction fragment sizes of 2 kbp and 250 bp from the *Eco*RI riboprint are very similar to the expected *Eco*RI fragment sizes from the MAPSORT analysis for all three *A. polyphaga* ssrRNA sequences. Moreover, the estimated sizes of the observed restriction fragments for the *Alu*I riboprint also closely resemble the expected sizes of the five largest *Alu*I digestion fragments for *A. polyphaga* strain BCM:0173:16 ATCC 50371. The size of the complete PCR product and the similarity of the *Eco*RI and *Alu*I riboprints to known *A. polyphaga* ssrRNA sequences, confirms that the 2.25 kbp amplicon is the *A. polyphaga* nuclear ssrRNA gene.

### **Optimisation of the PCR conditions**

Initial attempts to amplify the ssrRNA from strains of *Vannella* and *Platyamoeba* isolates were variable in success. It was evident that amplification of some strains was straight forward while others presented considerable difficulty. To optimise amplification of ssrRNA genes the PCR efficiency was examined with respect to the template DNA, PCR stringency, DNA polymerase and other PCR reagents.

### **Template DNA**

Examining the PCR efficiency from several DNA extractions allowed some conclusions to be drawn about the preparation of template DNA. It was apparent that the presence of bacteria in the extracted amoebal culture did not effect PCR efficiency, although bacterial material rendered estimation of the concentration of amoebal DNA more difficult. Therefore, as many bacteria were removed from cultures as possible, prior to DNA extraction, and template DNA was serially diluted to determine the optimal concentration of DNA for each extract. The efficiency of the PCR did not differ between the proteinase K and SDS methods of DNA extraction.

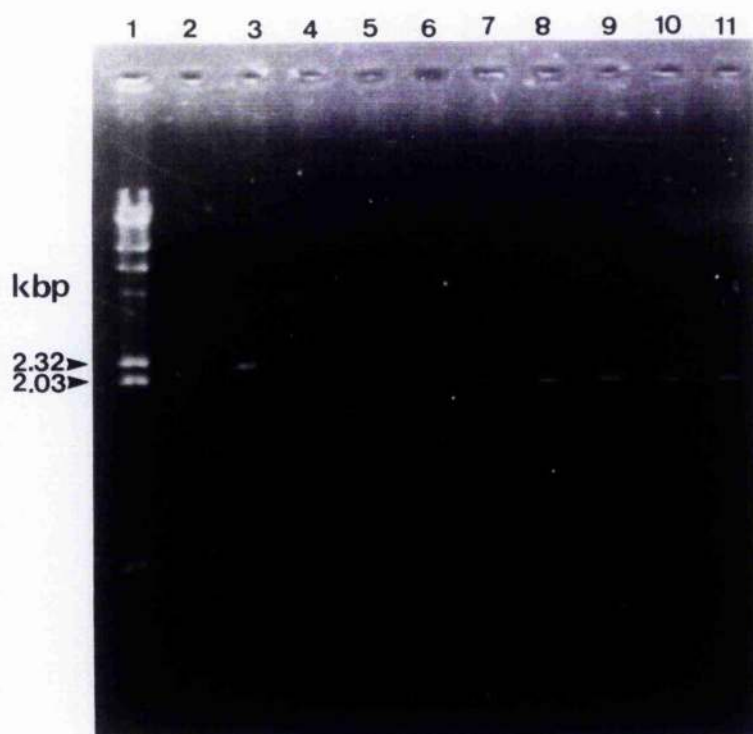


The use of ribonuclease treatment to remove RNA from DNA preparations did not affect PCR efficiency, but amplification of DNA preparations with other impurities was difficult. The purity of a DNA extraction was measured by the optical density ratio at wavelengths of 260 and 280 nm using spectrophotometry. Ratios of less than 1.75 indicated relatively impure DNA, and successful amplification from these preparations was less likely. The source of impurity was likely to be carbohydrate or small insoluble particles of cellular debris which were carried over following centrifugation and remained in the aqueous phase during purification with organic solvents. Template DNA was therefore prepared from dense, healthy cultures of amoebae after removal of the majority of bacteria. Only DNA preparations which produced a high molecular weight band when examined by agarose gel electrophoresis with a OD 260:280 ratio above 1.75 were used. Ribonuclease treatment was not used. The optimum DNA template concentration for PCR had to be determined for each extraction by serial dilution. The optimal amount of amoebal DNA template was between 10 and 150 ng.

### PCR stringency

The stringency of the PCR protocol is of paramount importance in determining the successful amplification of the correct target sequence. For a given set of primers, the stringency is chiefly determined by the magnesium chloride concentration and the annealing temperature. The optimum magnesium chloride concentration for the amplification of the *ssrRNA* genes from two strains of *Vannella* was determined by varying the final concentration from 0.5 mM to 4.5 mM. For *V. anglica* there was no amplification with concentrations up to 1.5 mM, poor amplification at 2.0 mM and improved amplification from 2.5 mM to 4.0 mM (Figure 2.6).

Similar results were obtained for *V. septirentionalis* where there was poor amplification at 1.5 mM, improved amplification at 2.5 mM and 3.5 mM and no amplification at a higher concentration of 4.5 mM magnesium chloride. These results demonstrate that the optimum magnesium chloride concentration for the amplification of the *ssrRNA* from these strains lies between 2.5 mM and 3.5 mM. Considering that the fidelity of *Taq* DNA replication is reduced at higher magnesium chloride concentrations, a final concentration of 2.5 mM was used in all subsequent amplifications.



**Figure 2.6** Effect of magnesium chloride concentration on PCR amplification.

Lane 1:  $\lambda$  HindIII

Lane 2: No DNA (negative control)

Lane 3: *Acanthamoeba polyphaga* (positive control)

Lane 4: *Vannella anglica* (0.5 mM  $\text{MgCl}_2$ )

Lane 5: *V. anglica* (1.0 mM  $\text{MgCl}_2$ )

Lane 6: *V. anglica* (1.5 mM  $\text{MgCl}_2$ )

Lane 6: *V. anglica* (1.5 mM  $\text{MgCl}_2$ )

Lane 7: *V. anglica* (2.0 mM  $\text{MgCl}_2$ )

Lane 8: *V. anglica* (2.5 mM  $\text{MgCl}_2$ )

Lane 9: *V. anglica* (3.0 mM  $\text{MgCl}_2$ )

Lane 10: *V. anglica* (3.5 mM  $\text{MgCl}_2$ )

Lane 11: *V. anglica* (4.0 mM  $\text{MgCl}_2$ )

A series of PCRs were used to identify the optimal annealing temperature for the amplification of *ssrRNA* genes. It was soon evident, however, that all strains would not successfully amplify at the same annealing temperature. Table 2.4 shows which strains were capable of amplification over several annealing temperatures ranging from 37°C to 52°C. The table adopts a scheme where the yield of the PCR product, determined by arbitrarily quantifying the fluorescence of the PCR product on an ethidium bromide stained agarose gel, is represented by the number of asterisks; three for a good yield, two for a moderate yield and one for a poor yield of product. A dash represents no amplification. From the table, a range of optimal annealing temperatures are apparent and it is clear that the yields of PCR product varied markedly between the different isolates. Isolates of *A. polyphaga*, *V. platypodia*, *V. septentrionalis* and *P. calycinucleolus* produced good yields of PCR products over a wide range of annealing temperatures. Some isolates, such as *V. simplex* and *V. lata*, required more specific annealing temperatures, whereas other isolates including *V. devonica*, *V. anglica* and *P. plurinucleolus* produced generally poor amplification products over the range of temperatures. In summary, the greatest yields of PCR products, for most isolates, were obtained with an annealing temperature of 46°C or 49°C.

Polymerase Match Enhancer (PME) was also included in the original protocol to improve the stringency of the PCR. Although as little as 0.25 units of PME could prevent low molecular weight smearing, PME also prevented amplification of *ssrRNA* genes which were otherwise produced in low yields. Consequently, PME was not used in subsequent amplifications.

### **DNA polymerases**

Increasing the amount of *Taq* DNA polymerase (Perkin Elmer) beyond 1 unit did not significantly improve the yield of PCR product per reaction. Indeed, with a new batch of enzyme, as little as 0.05 units was sufficient to amplify the *ssrRNA* gene. The efficiency of two other *Taq* DNA polymerases with their recommended buffers, from Gibco BRL and NBS Biologicals, was also examined. The Gibco BRL *Taq* DNA polymerase performed similarly to the Perkin Elmer enzyme. The *Taq* DNA polymerase from NBS biologicals was a considerably cheaper enzyme, which on occasion produced higher yields of PCR products, but the performance was generally less consistent. This enzyme often produced low molecular weight smearing, which was presumably due to either premature dissociation of the enzyme from the DNA

**Table 2.3** Optimal annealing temperature for PCR amplification of ssrRNA genes.

Amoeba Isolate	PCR Annealing Temperature					
	37°C	40°C	43°C	46°C	49°C	52°C
<i>Acanthamoeba polyphaga</i>	**	.	**	**	***	***
<i>Vannella platypodia</i>			***	***	***	
<i>V. simplex</i>		-	-	-	*	*
<i>V. devonica</i>			*	**	*	*
<i>V. arabica</i>			*	-	*	-
<i>V. anglica</i> 1589/8			**	*	**	**
<i>V. septrentionalis</i>			-	***	***	***
<i>V. anglica</i> 1589/11			*	*	*	*
<i>V. lata</i>			-	-	***	
<i>Platyamoeba mainensis</i>			*	**	***	
<i>P. placida</i>			*	**	**	-
<i>P. stenopodia</i>		-	-	*	-	-
<i>P. calycinucleolus</i>			*	**	***	***
<i>P. plurinucleolus</i> 1565/7			-	*	**	*
<i>P. australis</i>			-	**	*	
<i>P. bursella</i>	**	*	**	*	*	-
<i>P. plurinucleolus</i> 1565/11			*	*	**	
VP3				***	***	-
Species E					***	***
SIA					*	*

The symbols \*\*\*, \*\*, \*, - represent good, moderate, poor and no ssrRNA PCR amplification respectively.

template, poor enzyme stability or DNA nuclease activity. However, this problem could usually be overcome by the addition of bovine serum albumin to the reaction mixture. The enzyme activity of the NBS *Taq* DNA polymerase also varied considerably from batch to batch. Because of the inconsistencies with this enzyme, the more expensive *Taq* DNA polymerase from Perkin Elmer was used in subsequent reactions.

### **Other PCR reagents**

The addition of 10% glycerol or 0.1% to 10% Nonidet P-40 detergent did not improve the efficiency of PCR, hence, these reagents were not included in the final PCR conditions. The use of freshly mixed dNTPs was recommended, however, as repeated freeze-thawing of a dNTP mix slowly reduced the PCR efficiency over time. The effect of altering the primer and dNTPs concentrations was not investigated adequately to warrant constructive comment.

### **Inconsistency of PCR amplification**

Despite considerable effort to optimise the PCR for amplification of *ssrRNA* genes from amoebae, amplification was problematic on occasion for many of the isolates, resulting in variable yields of amplification products. Table 2.4 provides a summary of the difficulty involved with consistent amplification for each of the strains. Those strains which produced good yields of amplicons were the most consistently amplified, whereas the strains which produced low yields were the most difficult to amplify, regardless of the optimised protocol. The most difficult isolate was *P. stenopodia* which was never satisfactorily amplified despite using at least ten independent DNA preparations.

### **Final PCR conditions**

Amplification of amoeba *ssrRNA* genes was carried out in a 500 µl Eppendorf tube containing 2 µl of amoebal template DNA (10-100 ng), 5 µl of 10x PCR buffer (100 mM Tris-HCl, pH 8.3, 500 mM KCl), 250 ng of bovine serum albumin, 25 µM of each ultra-pure dNTP, 20 pmol of each universal eukaryotic *ssrRNA* primer and 2.5 mM MgCl<sub>2</sub> and sterile ultra-pure water (Milli-Q Plus) to a volume of 48 µl. The reaction mixture was overlaid with 100 µl of light mineral oil, preheated at 95°C for 5 min on a Minicycler Model PTC-150 thermocycler and cooled to 72°C before adding 2 µl (1 unit) *Taq* DNA polymerase (Perkin Elmer). The PCR mixture was subjected to 30 cycles of denaturing at 94°C for 1.5 min, annealing at 46°C or 49°C

for 1.5 min, a stepped increase of 1°C every 5 sec to 72°C and 2 min at the elongation temperature of 72°C. The polymerase chain reaction was then completed with a final 5 min incubation at 72°C to ensure that all polymerization was completed.

### **Amplification of ssrRNA genes from fan-shaped amoebae**

Successful PCR amplifications of the presumptive ssrRNA from *Vannella*, *Platyamoeba* and unknown isolates were examined by agarose gel electrophoresis stained with ethidium bromide (Figures 2.7 A, B and C). The estimated sizes of the amplicons were determined by comparing products with a  $\lambda$  HindIII DNA marker. All strains of *Vannella*, except *V. platypodia*, produced a single amplification product of approximately 1.95 kbp (Figure 2.7 A). Minor differences in size may be present between these strains but were not clearly resolved under the electrophoresis conditions used. The size of the *V. platypodia* amplicon was approximately 200 bp smaller than the other isolates at approximately 1.75 kbp (Lane 2). The size of the presumptive ssrRNA PCR products for the *Platyamoeba* were again conserved at approximately 1.95 kbp (Figure 2.7 B). There was no observable difference in the size of the *Platyamoeba* amplicons. One strain, *P. stenopodia*, was particularly difficult to amplify. On the only successful amplification of *P. stenopodia* a weak 1.95 kbp product and a secondary product of 1.5 kbp were produced. Amplification of a *P. bursella* product was difficult at annealing temperatures of 46°C and 49°C. At lower temperatures the production of a 1.95 kbp amplicon coincided with secondary products; one of 1.5 kbp, and a more prominent product of 1.2 kbp. Successful amplifications of the unidentified isolates, including the *Vannella* sp. ssrRNA gene from the recombinant pBluescript plasmid, are shown in Figure 2.7 C. The *Vannella* sp. ssrRNA gene amplification product was approximately 1.9 kbp in size (Lane 2), consistent with the *Vannella* and *Platyamoeba* amplifications. Species E and SIA both produced amplicons of approximately 1.95 kbp, although the yield of the SIA product was very poor (Lanes 4 and 5). The unidentified VP3 isolate produced a marginally larger product of 2.0 kbp (Lane 3). Together all the *Vannella*, *Platyamoeba* and unidentified morphologically similar isolates produced presumptive ssrRNA gene amplifications of between 1.9 kbp and 2.0 kbp, with the notable exception of *V. platypodia*, which produced a substantially smaller product of about 1.75 kbp.

**Figure 2.7** Small-subunit ribosomal RNA PCR amplifications from fan-shaped amoebae. (A) *Vannella* species, (B) *Platyamoeba* species and (C) unidentified isolates.

**A.**

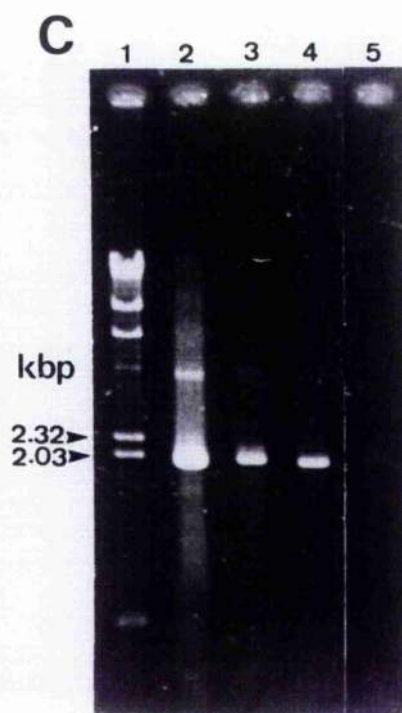
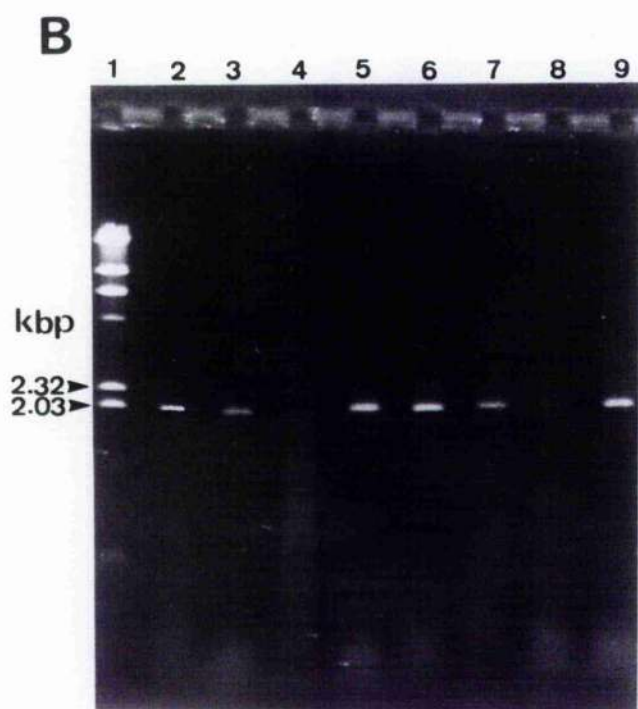
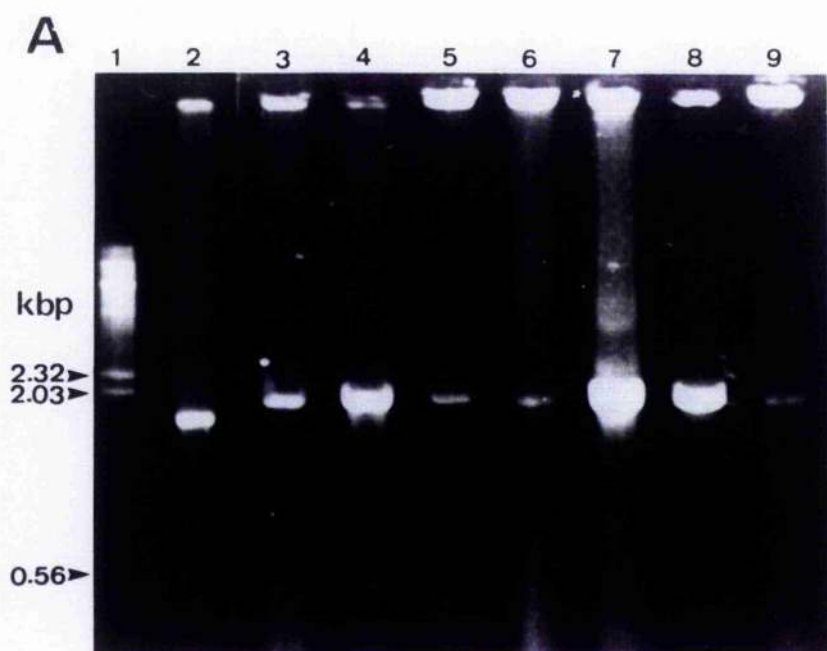
Lane 1:  $\lambda$  HindIII  
Lane 2: *Vannella platypodia* CCAP 1589/2  
Lane 3: *V. simplex* CCAP 1589/3  
Lane 4: *V. devonica* CCAP 1589/5  
Lane 5: *V. arabia* CCAP 1589/7  
Lane 6: *V. anglica* CCAP 1589/8  
Lane 7: *V. septrentionalis* CCAP 1589/10  
Lane 8: *Vannella* sp. CCAP 1589/11  
Lane 9: *V. lata* CCAP 1589/12

**B.**

Lane 1:  $\lambda$  HindIII  
Lane 2: *Platyamoeba mainensis* CCAP 1565/1  
Lane 3: *P. placida* CCAP 1565/2  
Lane 4: *P. stenopodia* CCAP 1565/3  
Lane 5: *P. calycinucleolus* CCAP 1565/6  
Lane 6: *P. plurinucleolus* CCAP 1565/7  
Lane 7: *P. australis* CCAP 1565/9  
Lane 8: *P. bursella* CCAP 1565/10  
Lane 9: *P. plurinucleolus* CCAP 1565/11

**C.**

Lane 1:  $\lambda$  HindIII  
Lane 2: pBluescript *Vannella* sp. F-49  
Lane 3: VP3  
Lane 4: Species E VP14  
Lane 5: SIA





Due to the inconsistency in PCR amplification and the use of two different annealing temperatures, the amoebal amplifications shown in Figure 2.7 were not obtained on a single occasion, and the negative control reactions are not shown. Negative control reactions, with either no template DNA or bacterial DNA, were run alongside the experimental reactions, but at no stage did the negative control produce an amplification product.

### **Confirmation of amoeba *ssrRNA* amplifications by hybridization**

To provide further evidence for the successful amplification of amoebae *ssrRNA* genes, a hybridization experiment was performed. A recombinant plasmid, which contained the entire nuclear *ssrRNA* gene from *Vannella* sp. F-49 ATCC 30945, was obtained from Dr M.L. Sogin, Woods Hole, Massachusetts, USA. The gene was amplified from the plasmid using the universal primers, as previously described, and a *ssrRNA* probe was produced using random hexanucleotide primers and a digoxigenin DNA labelling kit.

The probe was hybridized to the recombinant plasmid and bacterial *ssrRNA* amplicon to determine the specificity of the *ssrRNA* probe, and all presumptive *Vannella* *ssrRNAs* to confirm that the amplifications were derived from nuclear *ssrRNA* genes of amoebae. Recombinant plasmids and PCR amplicons were examined by ethidium bromide stained agarose gel electrophoresis, blotted on to a nylon membrane, hybridized with the DIG-labelled probe and exposed to X-ray film. Photographs of the ethidium bromide stained gels and autoradiographs, for the probe specificity and the presumptive *Vannella* *ssrRNA* hybridizations are shown in Figures 2.8 and 2.9 respectively.

### **Specificity of the digoxigenin-labelled ribosomal probe**

By comparing the photographs of the ethidium bromide agarose gel and the corresponding image of the blot, it is clear that the ribosomal probe hybridized strongly to the positive controls (Figure 2.8 A and B); the linearised recombinant plasmid, which contains the ribosomal probe template (Lane 2), and the 1.9 kbp PCR-amplified *Vannella* sp. *ssrRNA* probe template (Lane 4). The ribosomal probe, however, also hybridized to linearised non-recombinant plasmid (Lane 3). This indicates that the ribosomal probe includes some labelled plasmid DNA. Plasmid DNA may have been carried over from the amplification of the *Vannella* sp. *ssrRNA* gene, and also labelled.

Eukaryotic and eubacterial primers were used to amplify the *ssrRNA* genes from DNA extracted from a culture of *V. devonica* containing bacteria, and a culture of bacteria alone. Using the eukaryotic primers, the *V. devonica* DNA produced the expected 1.95 kbp PCR product, which hybridized strongly with the ribosomal probe (Lane 5). No visible product was detected when these primers were used to amplify *ssrRNA* from the bacterial DNA, and there was only a weak smear of hybridization which is presumably due to the probe adhering to the eukaryotic primers (Lane 6). This demonstrates that the bacterial *ssrRNA* gene could not be amplified using eukaryotic rRNA primers.

Using the eubacterial primers, amplifications from both the bacterial DNA extraction and the *V. devonica* DNA extraction (which includes bacterial DNA), produced a product of 1.5 kbp (Lanes 7 A and 8 A). A product of this size is consistent with the size of the *Escherichia coli* *ssrRNA* gene (Brosius *et al.*, 1978). However, the *Vannella* *ssrRNA* probe did not hybridize to these bacterial *ssrRNA* amplification products (Lanes 7 B and 8 B). Despite reducing the stringency of the reaction, by lowering the hybridization temperature to 60°C, the *Vannella* *ssrRNA* probe failed to hybridize to the eubacteria *ssrRNA* PCR amplifications. While acknowledging that the *Vannella* *ssrRNA* probe must contain some labelled probe specific to plasmid DNA, these results demonstrate that the probe is able to discriminate between eubacteria and eukaryotic *ssrRNA* gene amplifications, by specifically hybridizing to eukaryotic *ssrRNA* genes.

### Hybridization with *Vannella* *ssrRNA* PCR amplifications

Presumptive *ssrRNA* PCR products from eight strains of *Vannella* encompassing seven species were examined on an ethidium bromide stained agarose gel (Figure 2.9 A). The size of the presumptive *ssrRNA* PCR products was consistently 1.95 kbp (Lanes 3-9) for all strains of *Vannella* except *V. platypodia* which produced a smaller product of 1.75 kbp (Lane 2). The DNA samples were blotted from the gel on to a nylon membrane, hybridized with the digoxigenin labelled eukaryotic specific *ssrRNA* probe. The subsequent image is shown in Figure 2.9 B. All the presumptive *Vannella* *ssrRNA* amplicons hybridized strongly with the probe, although some stained very faintly with ethidium bromide. Considering the specificity of the *ssrRNA* probe and that there was no plasmid DNA or any eukaryotic DNA other than amoebae DNA present, the strong hybridization confirms that these amplifications were derived from *Vannella* nuclear *ssrRNA* genes. The similar size

**Figure 2.8** Small-subunit ribosomal RNA probe hybridization specificity.

(A) Photograph of the ethidium bromide stained agarose gel.

(B) Photograph of the autoradiograph.

Lane 1, 9:  $\lambda$  HindIII

Lane 2: pBluescript *Vannella* sp ssrRNA EcoRV

Lane 3: pBluescript EcoRV

Lane 4: *Vannella* sp F-49 ATCC 30945 ssrRNA

Lane 5: *V.devonica* amplification with eukaryotic ssrRNA primers

Lane 6: Bacterial amplification with eukaryotic ssrRNA primers

Lane 7: *V.devonica* amplification with eubacteria ssrRNA primers

Lane 8: Bacterial amplification with eubacterial ssrRNA primers

**Figure 2.9** Hybridization of the ssrRNA probe with *Vannella* amplifications.

(A) Photograph of an ethidium bromide stained agarose gel.

(B) Photograph of the autoradiograph.

Lane 1:  $\lambda$  HindIII

Lane 2: *Vannella platypodia* CCAP 1589/2 ssrRNA amplification

Lane 3: *V. simplex* CCAP 1589/3 ssrRNA amplification

Lane 4: *V. devonica* CCAP 1589/5 ssrRNA amplification

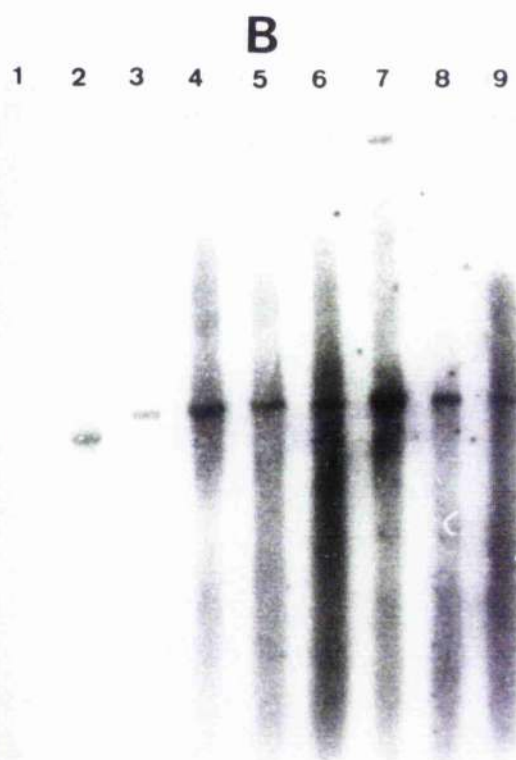
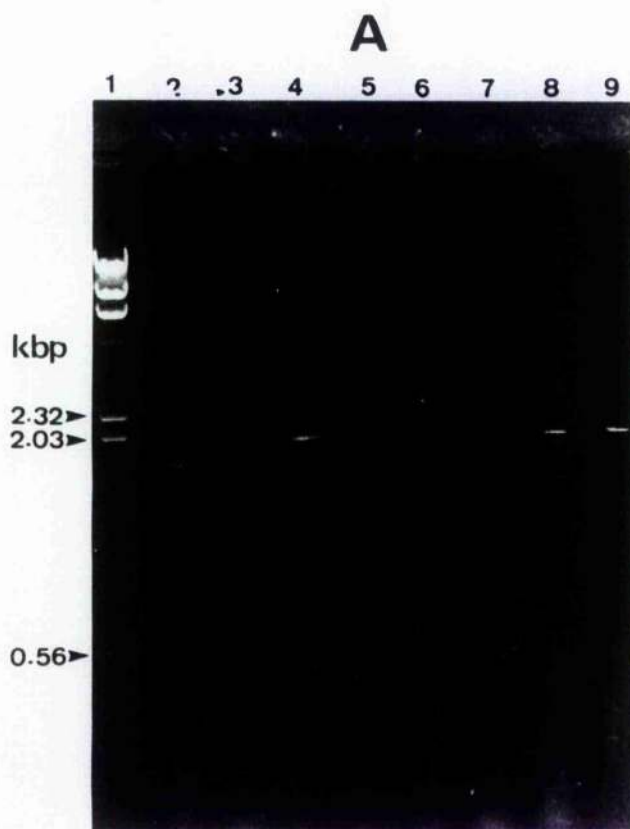
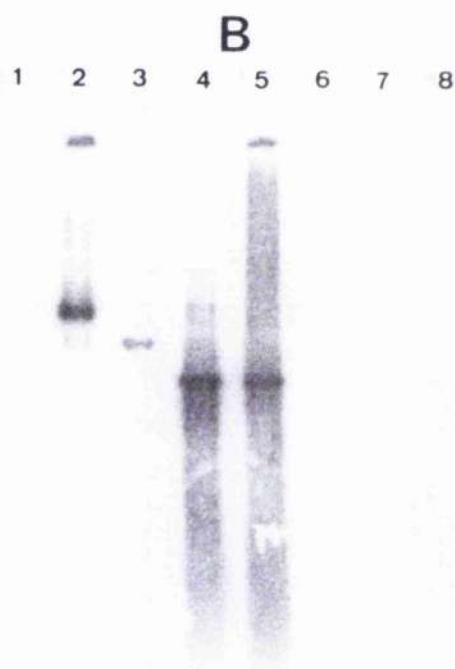
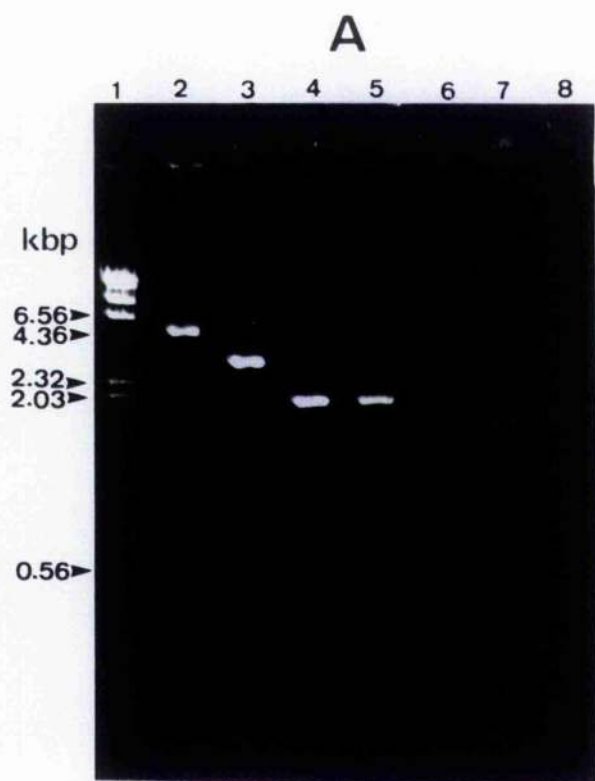
Lane 5: *V. arabia* CCAP 1589/7 ssrRNA amplification

Lane 6: *V. anglica* CCAP 1589/8 ssrRNA amplification

Lane 7: *V. septrentionalis* CCAP 1589/10 ssrRNA amplification

Lane 8: *Vannella* sp. CCAP 1589/11 ssrRNA amplification

Lane 9: *V. lata* CCAP 1589/12 ssrRNA amplification



of the amplicons from *Platyamoeba* and the unidentified isolates infers that these products were also derived from nuclear *ssrRNA* genes of amoebae.

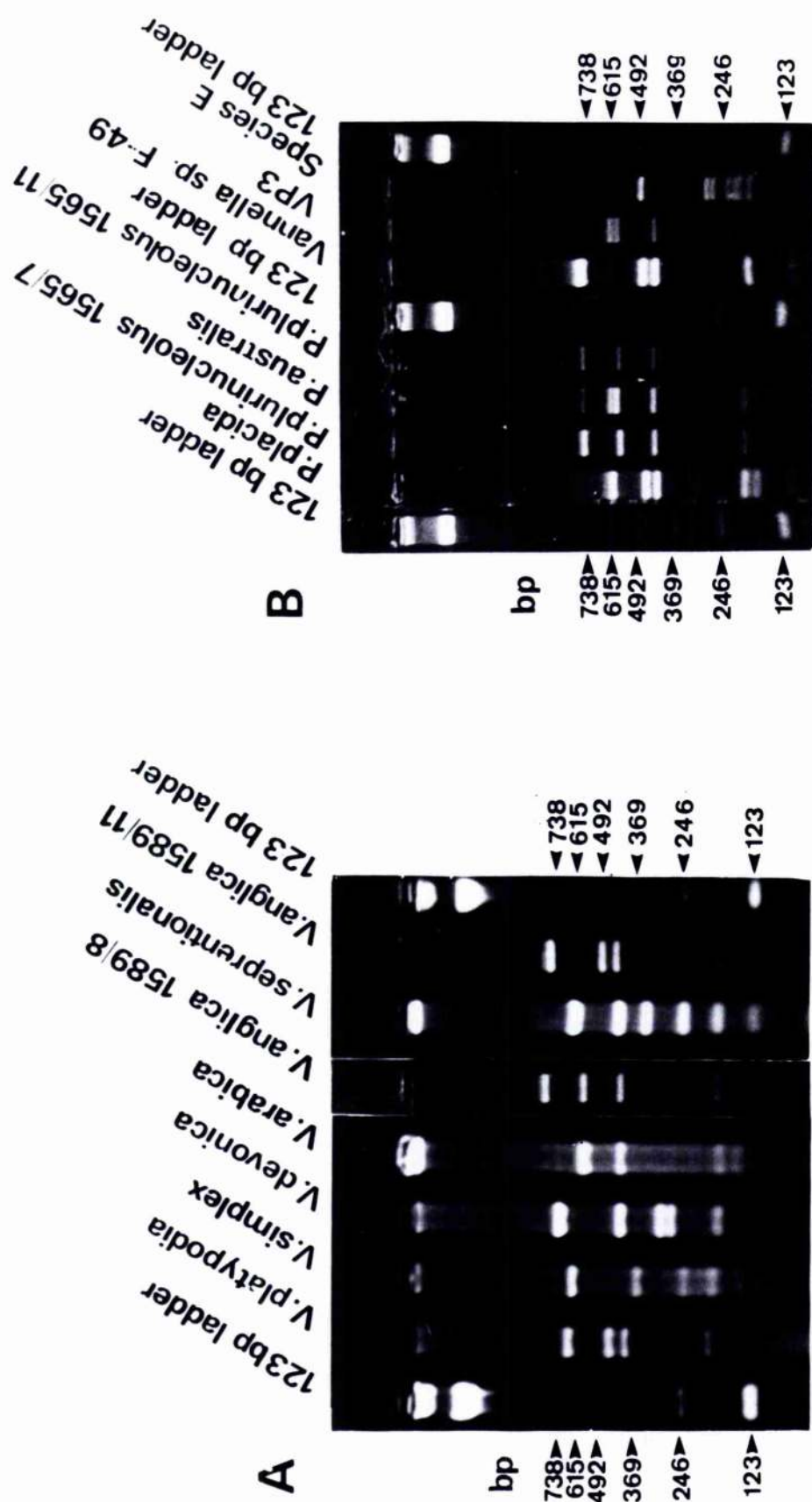
### Restriction analysis of amoebae *ssrRNA* genes

It was an intention to use *ssrRNA* gene restriction analysis (riboprinting) for two purposes; to characterize *ssrRNAs* to assess the utility of this method for identifying fan-shaped amoebae isolated from the field, and to examine the phylogenetic relationships among these morphologically similar isolates. Unfortunately, due to difficulties amplifying high yields of *ssrRNA* amplicons it was not possible to carry out restriction analyses on all the isolates using a wide range of restriction enzymes. To characterize *Vannella* and *Platyamoeba* species, riboprinting was carried out on 14 isolates with *AluI* and 15 isolates with *Hinfi*.

The *ssrRNA* amplicons were digested with *AluI* and *Hinfi* restriction enzymes and separated on a 2% agarose gel, stained with ethidium bromide exposed to ultraviolet light and photographed. Both enzymes produced an average of five restriction fragments. The *AluI* and *Hinfi* restriction digests are shown in Figures 2.10 and 2.11, respectively. Interpretations of the *AluI* and *Hinfi* riboprints, showing the estimated sizes of the digestion fragments and the total gene length are given in Tables 2.4 and 2.5.

### *AluI* riboprints

Examination of the *AluI* riboprints revealed substantial variation, particularly among the *Vannella* isolates (Figure 2.10). Only the riboprints of *V. anglica* 1589/11 and the unknown *Vannella* sp F-49 were identical, although they initially appeared to be similar to the riboprints of *V. platypodia* and *P. placida*. Closer examination of the sizes of the individual bands confirmed that four bands of *V. anglica* 1589/11 and *Vannella* sp. F-49 at approximately 480 bp, 425 bp 175 bp and 100 bp were shared with *P. placida*, but only one band, at approximately 175 bp, was present in *V. platypodia*. The *Platyamoeba* *AluI* riboprints were more similar to each other than the *Vannella* *AluI* riboprints. The riboprints of *P. plurinucleolus* strains 1565/7 and 1565/11, and *V. anglica* 1589/8 were identical. With a similar riboprint, VP3 and *P. australis* were also identical, accepting that the largest 780 bp fragment from *P. australis* is incompletely digested. Two shared *AluI* fragments were apparent in most of the isolates examined. One fragment ranged in size from 175 bp to 190 bp, perhaps marginally higher in *V. platypodia*, was present in all 14 strains, and a



**Figure 2.10** Restriction digests of ssrRNA PCR amplifications with *Alu*I. (A) *Vannella* species, (B) *Platyamoeba* species and unidentified isolates.

**Table 2.4** Representation of the ssrRNA *AluI* digestion fragments. The estimated sizes (bp) of the *AluI* fragments for each isolate are given in the columns with the estimated total gene length (kbp). Predicted homologous *AluI* fragments are presented in the same row.

	Species E										
	V. plautodis										
	V. simplex										
	V. devonica										
	V. arabica										
	V. anglica 1589/8										
	V. septentrionalis 1598/11										
	V. anglica 1598/11										
	P. placida										
	P. pleroneolus 1568/7										
	P. anstralis										
	P. pleroneolus 1568/11										
	Vannella sp. F-49										
	VP3										
A1											
A2	645	635			800	635		780		780	
A3						585			620		625
A4						580			570	580	600
A5						570					
A6											490
A7	450							480	470	470	
A8											
A9	395			415	415	415	420	425	425	430	425
A10		355									
A11				300			330				
A12				275							280
A13											250
A14		240					245				
A15											235
A16											215
A17											210
A18	195	180	175	175	180	175	175	190	190	185	185
A19		175						165	165		165
A20		145		145							
A21							120				
A22		110					100	105		100	
A23		100									
	1.63	1.94	1.90	1.88	1.98	1.92	1.96	1.97	1.96	1.97	1.87

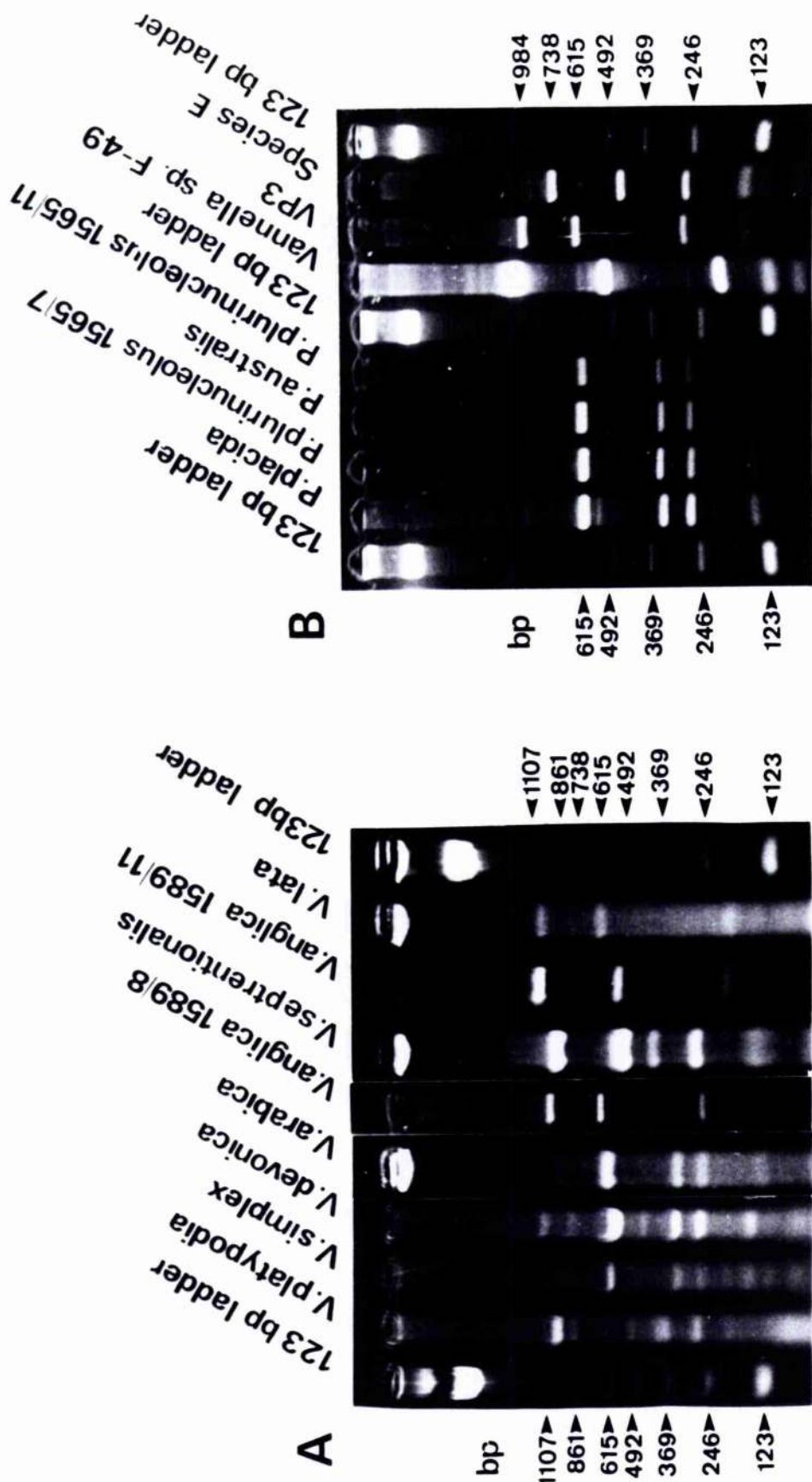
second band, ranging in size from 415 bp to 430 bp, was present in all strains except *V. platypodia*, *V. simplex* and Species E. Further examination of the Species E *AluI* riboprint suggests that it has the most distinctive restriction pattern with at least five unique bands.

The estimates of the total gene size, determined from the *AluI* fragments (Table 2.4), are generally in good agreement with the estimates from analysis of the complete gene amplifications of 1.75 kbp for *V. platypodia* and 1.95 kbp for the remaining isolates (Figure 2.7). The marginally smaller estimates of the total gene size for isolates of *V. platypodia*, *V. devonica*, *V. arabica*, *V. septrentionalis* and Species E, can be accounted for by small digestion fragments, smaller than 80 bp in size, which could not be detected on the riboprint. The largest estimate of total gene length using *AluI* fragments was 2.0 kbp for VP3, slightly higher than other isolates, but nevertheless an increase which was also detected from analysis of the complete gene amplifications previously described.

### ***HinfI* riboprints**

As with the *AluI* riboprints, there was considerably more variation in the *HinfI* restriction patterns within the *Vannella* isolates than within *Platyamoeba* isolates (Figure 2.11). In fact, all four *Platyamoeba* isolates had identical *HinfI* riboprints with fragments at approximately 135 bp, 260 bp, 340 bp and co-migrating fragments at approximately 610 bp. This riboprint motif was also present for *V. arabica*, and similar restriction patterns were found for *V. simplex* and *V. devonica*. In the case of *V. simplex*, the band of approximately 630 bp was of lower intensity than the *Platyamoeba* isolates, and additional fragments were found between 90 and 220 bp for the *HinfI* riboprint. This suggests that *V. simplex* has additional restriction sites within a region equivalent to one of the co-migrating *HinfI* fragments in the *Platyamoeba* isolates. The *V. devonica* *HinfI* riboprint also shows a close resemblance to the *Platyamoeba* *HinfI* riboprints, but due to incomplete digestion of the *ssrRNA* gene it is not clear whether or not one of the co-migrating fragments was digested further into smaller fragments. The *V. anglica* 1589/11 and *Vannella* sp. F-49 isolates had identical *HinfI* riboprints and were similar to *V. lata*. Isolates *V. anglica* 1589/8 and VP3 also had identical riboprints. Of the remaining isolates not previously mentioned, *V. platypodia* produced bands at 275 bp, 375 bp and 1000 bp. Other bands at 385 bp and 90 bp may also be present but incomplete digestion made interpretation difficult. Finally, *V. septrentionalis* and Species E again had





**Figure 2.11** Restriction digests of ssrRNA PCR amplifications with *Hin*FI. (A) *Vannella* species, (B) *Platyamoeba* species and unidentified isolates.

**Table 2.5** Representation of the ssrRNA *Hinf*I digestion fragments. The estimated sizes (bp) of the *Hinf*I fragments for each isolate are given in the columns with the estimated total gene length (kbp). Predicted homologous *Hinf*I fragments are presented in the same row.

	<i>V. platypodia</i>	<i>V. simplex</i>	<i>V. devonica</i>	<i>V. arabica</i>	<i>V. anglica</i> 1589/8	<i>V. septentrionalis</i> 1598/1	<i>V. lata</i>	<i>P. placida</i>	<i>P. plurinucleolus</i> 1565/7	<i>P. australis</i> 1565/11	<i>P. plurinucleolus</i> sp. F-49	<i>VP3</i>	Species E
H1	1000					1040	1040				1000		
H2					950							970	
H3						860							
H4													745
H5		630	605	615	630		620	610	610	610		620	
H6			600?	610				605	605	605			
H7						540					520		
H8						510							
H9													445
H10	385?												
H11	375												
H12		345	345	340				325	340	340			
H13	275	275	275	270	265	275		260	260	260		270	260
H14		200				190	180				200		
H15		145	150			145							140
H16				140	140	135	135	130	135	135		140	140
H17		120	120?			125					120		115
H18	90?	90	90?			100					100		90
	?	1.80	?	1.97	1.98	1.92	1.99	1.97	1.95	1.95	1.94	2.00	1.93

very distinctive riboprints. The *V. septrentionalis* had unique bands at 860 bp and 510 bp (the band at 420 bp was clearly a incomplete digestion product), and Species E had unique bands at 745 bp and 445 bp. There were no *HinfI* restriction fragments which were common to all isolates.

Estimates of the total gene lengths using *HinfI* fragments (Table 2.5) were again in good agreement with the sizes determined from the complete gene amplification. Only the *V. simplex* *HinfI* estimate of 1.805 kbp deviated significantly from the expected size of 1.95 kbp for the complete gene. The difference was probably due to a small co-migrating *HinfI* fragment not being detected. However, from restriction analysis there does appear to be a small amount of variation in the total length of the *ssrRNA* gene. Excepting the obvious difference in size of *V. platypodia*, these amoebae *ssrRNA* genes may vary in length from 1.9 kbp to 2.0 kbp.

### Use of riboprinting for identification of field isolates

To identify unambiguously amoebae isolated from the field using riboprinting, it is essential that the technique unambiguously distinguishes between different species. In combination, the *AluI* and *HinfI* restriction enzymes were almost able to differentiate between the 12 CCAP reference strains examined. Only one pair of reference strains, *P. plurinucleolus* 1565/7 and 1589/11, had identical riboprints for both enzymes. The restriction fragment polymorphisms between species demonstrates that these enzymes can detect interspecific variation. The identical riboprints of both *P. plurinucleolus* strains also suggests that these enzymes cannot detect intraspecific variation, however, differences were apparent between *V. anglica* strains. The *Vannella* sp. F-49 *ssrRNA* produced identical *AluI* and *HinfI* riboprints to *V. anglica* 1589/11. Considering the interspecific variation observed within the group of amoebae, this indicates that these isolates may belong to the same species. Morphological observations on *Vannella* sp. F-49, however, suggest this is not the case (Nerad & Sawyer, pers. comm.).

From the restriction data it is not possible to determine unambiguously the identity of isolate VP3: isolate VP3 does not share both restriction patterns with any of the reference strains. The VP3 *AluI* restriction pattern is identical to *P. australis*, and is similar to *V. anglica* 1589/8, *P. placida* and both strains of *P. plurinucleolus*. The VP3 *HinfI* restriction pattern is identical to *V. anglica* 1589/8, but not *V. anglica* 1589/11. This riboprint is also similar to *V. arabica*, and the *Platyamoeba* strains.

In combination, the similarities between the riboprints of VP3 and other *Vannella* and *Platyamoeba* species indicate that VP3 is closely related to these amoebae, but there is sufficient differences to suggest that it may be an undescribed species.

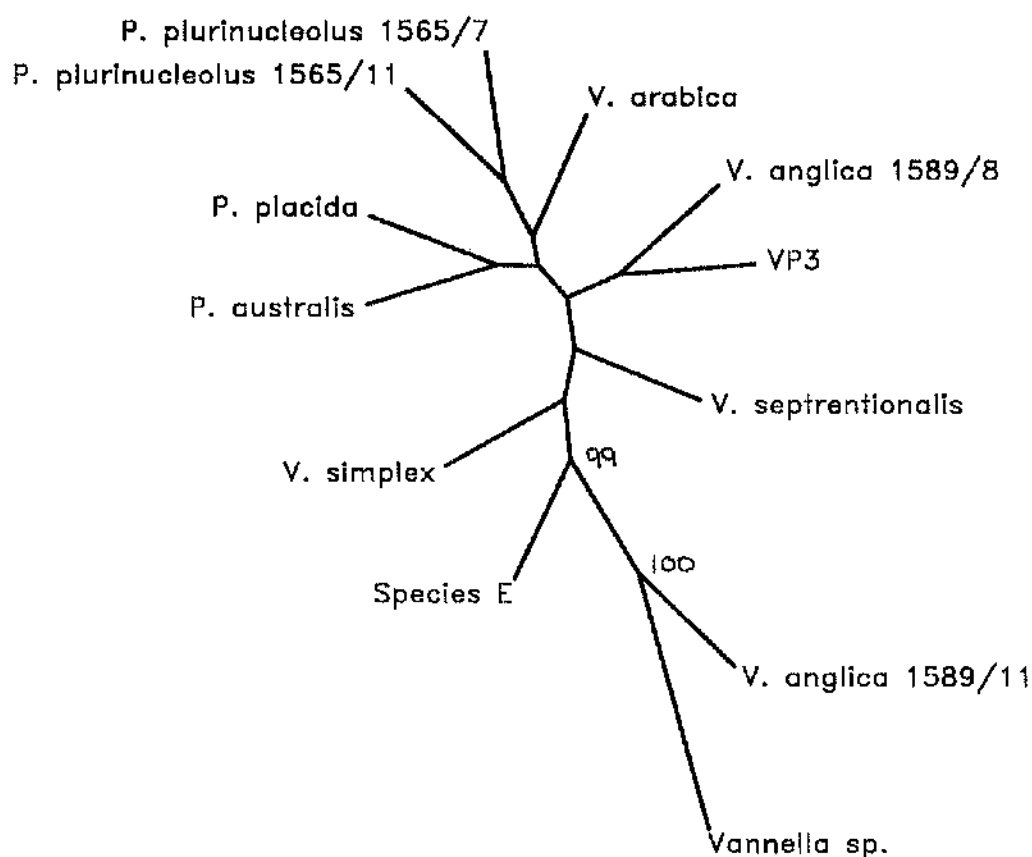
All the riboprints between the reference strains, except perhaps *V. platypodia* which was complicated by a significantly smaller gene size, clearly had some similarity. However, Species E was distinctly different from all the reference strains for both riboprints. Species E possessed five unique restriction fragments for the *AluI* digest, and at least two for the *HinfI* digest. Considering that the two largest *HinfI* fragments were unique to Species E, it is most likely that a majority of the smaller *HinfI* fragments are not in fact homologous to fragments of the same size in other strains. The riboprints of Species E therefore clearly demonstrate that this strain does not belong to any of the reference species, and the dissimilarity indicates that Species E may not be a *Vannella* or *Platyamoeba*.

### **Phylogenetic analysis of riboprinting data**

Restriction data was converted to a binary data set for parsimony analysis. Homologous restriction fragments from the *AluI* and *HinfI* digests were aligned in rows and each fragment was assigned a letter and a number (Tables 2.4 and 2.5). The binary sequence for each isolate was determined, where '1' signifies the presence and '0' the absence of each of the restriction fragments. The binary data set for the parsimony analysis is shown in Table 2.6. Data from three species was not included in the analysis. The substantial difference in size of the total ssrRNA of *V. platypodia*, compared to the other isolates, made it difficult to determine whether the equivalent restriction fragments were present in other isolates. For the other two strains, no *AluI* data was available for *V. lata*, and the *V. devonica* the *HinfI* restriction data was unclear due to partial digestion. Restriction fragments designated as being unique to these three isolates namely, A07, A09 and H10 were therefore excluded. The binary data was bootstrap resampled 100 times, and the most parsimonious tree for the amoebae ssrRNA restriction data is shown in Figure 2.12.



The tree appears to form one large cluster consisting of all the *Platyamoeba* isolates, *V. arabica*, *V. anglica* 1589/8 and VP3, and a second cluster consisting of *V. anglica* and *Vannella* sp. F-49. These clusters are separated by the branching of three isolates which had the most unique restriction patterns; *V. septrentionalis*, *V. simplex* and Species E. However, the bootstrap values are generally very low. Only two nodes occur in more than 50% of the trees from the 100 resampled data sets, and the high bootstrap values for these nodes are likely to be an artifact of the tree searching method. Therefore, there is no support for the clusters, or the branching order of these isolates within this phylogeny, and the use of just two restriction enzymes was therefore clearly insufficient to resolve the branching order of these morphologically similar amoebae.



**Figure 2.12** Most parsimonious unrooted phylogram for fan-shaped amoebae using *ssrRNA* restriction data. Digestion fragments which were judged to be homologous were treated as independent characters and used in a MIX parsimony analysis. The bootstrap values at nodes represent the percentage occurrence of the branch in 100 resampled data sets. Values less than 50% are not shown.

## Discussion

### **Isolation of amoebal DNA**

The greatest yields of high molecular weight DNA were obtained with proteinase K and SDS methods. These methods have been previously used to extract DNA which was suitable for restriction analysis, hybridization and PCR amplification from Gram positive bacteria and Gram negative bacteria (Fuhrman *et al.*, 1988; Loeffelholz & Scholl, 1989), and eukaryotic organisms including the schizopyrenid *Naegleria* and the lobose, naked amoeba *Acanthamoeba* (McLaughlin *et al.*, 1988; Gast *et al.*, 1996). In this study, comparisons with other extraction methods indicated that proteinase K and boiling SDS extraction were more effective than several other methods for extraction of high molecular weight DNA from naked amoebae.

The use of guanidine thiocyanate has been an effective reagent for isolating DNA in instances where high nuclease activity is problematic. The strong protein denaturant effectively lysed cells, but, precipitation was ineffective. Propan-2-ol and ammonium acetate have been used to specifically harvest bacterial DNA (Owen & Borman, 1987; Pitcher *et al.*, 1989). In this study, however, centrifugation at 10,000 g was insufficient to separate the organic and aqueous phases properly with the large amounts of insoluble protein which were formed.

Chelex has been a useful reagent in DNA extraction because it chelates metal ions that may act as catalysts in the breakdown of DNA at high temperature (Singer-Sam *et al.*, 1989). It has been successfully used to isolate DNA suitable for amplification from forensic samples, bacteria and the schizopyrenid genus *Vahlkampfia* (Walsh *et al.*, 1991; De Lamballerie *et al.*, 1992; Brown & De Jonckheere, 1994). In this study, however, in the absence of SDS, the lytic properties of boiling Chelex were poor and DNA recoveries were low. Chelex does not necessarily have any lytic properties, although the alkalinity of the Chelex suspensions and exposure to 100°C result in disruption of cells (Singer-Sam *et al.*, 1989). Previous studies have shown that boiling cells with Chelex alone is insufficient to release DNA from sperm and gram positive bacteria, and treatment with proteinase K or SDS is required (Walsh *et al.*, 1991; De Lamballerie *et al.*, 1992). Lysis was improved with the inclusion of SDS, but the combination of the



substantial uncovered volume when using Chelex, and the DNA lost during purification manipulations account for the low yield.

Genomic DNA in the size range of 20-40 kbp was routinely extracted using the proteinase K and SDS methods indicating that the techniques did not shear the DNA into small fragments, and the extractions were suitable for the amplification of the 2 kbp *ssrRNA* genes. Moreover, considering no discrete bands were detected, the DNA extractions indicate that the ribosomal DNA repeat unit of *Vannella* and *Platyamoeba* is unlikely to be present as high copy number extracellular elements. The rDNA repeat of the amoeboid-flagellate *Naegleria* is located on approximately 4000 circular 14 kbp elements which constitute approximately 17.5% of the cellular DNA (Clark & Cross, 1987). Circular or linear extrachromosomal rDNA elements of up to 19 kb are also present in high copy number in ciliates such as *Stylonychia*, *Tetrahymena* and *Paramecium* (Karrer & Gall, 1976; Findlay & Gall, 1978; Steinbrück *et al.*, 1981). The ribosomal repeat unit of the naked amoebae *Acanthamoeba* is 12 kbp in size, although it has not been determined whether the rDNA is integrated into chromosomes or is mostly extrachromosomal (D'Alessio *et al.*, 1983). Assuming that the rDNA repeat unit of *Vannella* and *Platyamoeba* is of a similar size, high copy number of extrachromosomal elements would have been detected in the DNA extractions as a distinct DNA band below 20 kbp in size. However, if the repeat element was much larger like those of *Physarum* and *Dictyostelium* which are in excess of 30 kbp (Vogt & Braun, 1976; Cockburn *et al.*, 1978), the extracted DNA was sufficiently sheared that the extrachromosomal DNA would not have been detected. Nevertheless, the DNA extractions indicate that the *Vannella* and *Platyamoeba* rDNA repeat unit exists only within chromosomes, or that the copy number of extracellular rDNA is low.

### PCR development

Initially, amplification of *ssrRNA* genes from *Vannella* and *Platyamoeba* isolates using PCR was unsuccessful. Several factors usually associated with PCR difficulties were investigated. Aspects of the template DNA, PCR stringency and PCR cycling conditions were all examined, but it was eventually discovered that the key factor was the default ramping rate of the thermal cycler. For successful amplification of *ssrRNA*, the ramping rate from the annealing to the elongation temperature was reduced to a 1°C increase every five seconds. This result suggests that the primers were becoming dissociated from the template DNA as the temperature increased

from the annealing to the elongation temperature. This may also indicate that one or both primers is not necessarily complementary to the template sequence.

Despite optimising PCR conditions with respect to template DNA, magnesium chloride concentration and annealing temperature, the amplification of *ssrRNA* genes remained quite inconsistent for many of the strains. There was no definite correlation between PCR efficiency and the amoebae cultures. However, there was a tendency for PCR performance to be better from cultures which could maintain large numbers of active cells, such as *V. platypodia* and *V. septreptionalis*, than those such as *V. arabica* and *P. placida* which could not. Previously it has been shown that the copy number of extrachromosomal rDNA may vary significantly according to the growth phase in *Tetrahymena* (Engberg & Pearlman, 1972). In *Amoeba proteus*, the synthesis of late replicating DNA, which is presumably extrachromosomal rDNA, also depends on feeding and culture conditions (Spear & Prescott, 1980; Makhlin, 1981). If a majority of the rDNA is extrachromosomal in *Vannella* and *Platyamoeba*, and the copy number is significantly lower during unfavourable growth, this may account for the difficulty amplifying the *ssrRNA* gene from some strains.

*Platyamoeba stenopodia* was the most difficult isolate to amplify and was the only isolate to digest or burrow into the surface of the agar plate. This suggests that perhaps the agar itself may contain inhibitors of *Taq* DNA polymerase which may be carried through the DNA extraction process, although the purification steps incorporated into the extraction procedures should remove or disable potential inhibitors.

The poor amplification results may be due to degeneracy in the DNA sequences in the conserved regions at the ends of the gene to which the primers are targeted. The eukaryotic *ssrRNA* primers were originally designed from the analysis of 45 eukaryotic *ssrRNA* sequences (Medlin *et al.*, 1988). Since then, the extensive use of these, and similar primers for PCR amplification of the complete *ssrRNA* gene prior to DNA sequencing, has resulted in a dramatic increase in the number of eukaryotic *ssrRNA* sequences from 45 to over 620 (Van de Peer *et al.*, 1994). This has occurred without providing much information on the sequence composition of the 'conserved' sites at the termini of the gene, against which the primers are targeted.

### Amplification of amoeba ssrRNA genes

The first successful amplification was a 2.25 kbp product from the naked amoeba *Acanthamoeba polyphaga*. The size of the *A. polyphaga* amplification product was consistent with other ssrRNA genes examined in this genus, which are typically 2.3 kbp, among the largest ssrRNA genes known (Gunderson & Sogin, 1986; Gast *et al.*, 1994a, 1996). Comparing the restriction digestion patterns of the *A. polyphaga* amplification products with the expected restriction patterns derived from the complete ssrRNA sequences of three strains of *A. polyphaga* (Gast *et al.*, 1994b) on the GenEMBL database confirmed that the 2.25 kbp *A. polyphaga* amplification product was indeed the ssrRNA gene.

The majority of *Vannella* isolates produced an amplification product of approximately 1.95 kbp. The only exception was *V. platypodia* which produced a considerably smaller product of 1.75 kbp. To confirm that these amplification products were the ssrRNA genes, a probe was prepared from the ssrRNA gene from *Vannella* sp. F-49 ATCC 30945 by randomly labelling with digoxigenin. A plasmid containing the 1.9 kbp *Vannella* sp. ssrRNA was kindly donated by Dr M.L. Sogin. The probe did not hybridize with bacterial ssrRNA, but it hybridized strongly with all the *Vannella* isolates, demonstrating that the amplicons were derived from amoeba ssrRNA genes. Amplification from the morphologically similar *Platyamoeba* and unknown isolates all produced products of 1.9-2.0 kbp. The use of eukaryotic specific ssrRNA primers and the conservation in size of the amplicon demonstrates that the amplifications from the *Platyamoeba* and unknown isolates are also ssrRNA genes.

Previous reports have identified considerable variation in the size of the ssrRNA among amoebae and amoeboid-flagellates. The ssrRNAs of testate amoebae *Paulinella chromatophora* and *Euglypha rotunda*, and the naked amoeba *Hartmannella vermiformis* are 1.8 kbp in size, whereas in the majority of the amoeboid-flagellates they are 2.0-2.2 kbp (Gunderson *et al.*, 1994; Weekers *et al.*, 1994; Bhattacharya *et al.*, 1995).

Naked amoebae of the *Acanthamoeba* and *Phreatamoeba* genera have several large expansion segments in variable regions of the ssrRNA resulting in genes of 2.3 kbp and over 2.3 kbp, respectively (Gunderson & Sogin, 1986; Hinkle *et al.*, 1994). Group I introns are also a feature of two amoebae genera. Introns of 600-700 bp have been

found in *A. graffini* and *A. lenticulata* producing lengths of 2.9 and 3.0 kbp (Gast *et al.*, 1994a), and introns of 1.3 kbp have been found in *Naegleria* species which have genes of 3.3 kbp (De Jonckheere, 1993). The size of the *Vannella* and *Platyamoeba* *ssrRNA* genes is about average for eukaryotic *ssrRNAs*, and it is therefore unlikely that these amoebae possess introns. The absence or presence of introns will be confirmed by DNA sequence analysis of the complete *ssrRNA* gene from *V. anglica* (Chapter 3). The variation in size of the *ssrRNAs* among the *Vannella* and *Platyamoeba* isolates was presumably due to small insertions and deletions.

### **Characterization of amoeba isolates using riboprinting**

The *ssrRNA* genes of *Vannella*, *Platyamoeba* and unknown isolates were characterized by restriction analysis (riboprinting) to assess the usefulness of the technique for identification and phylogenetic analysis. A complete comparison of all *ssrRNA* genes was compromised by the low quantities of amplicons obtained for many of the isolates. Nevertheless, twelve reference strains covering ten species from the morphologically similar *Vannella* and *Platyamoeba* genera, and three unknown isolates were characterized using the *AluI* and *HinfI* restriction enzymes. In combination, almost every isolate had its own characteristic set of riboprints. Only the riboprints of *P. plurinucleolus* 1565/7 and *P. plurinucleolus* 1565/11, and the riboprints of *V. anglica* 1589/11 and *Vannella* sp. F-49 were identical for both the *AluI* and *HinfI* restriction enzymes. Since significant numbers of restriction polymorphisms were detected, despite the small proportion of the gene sequence sampled, these riboprints demonstrate that there is considerable sequence variation among these amoebae. Interspecific variation was detected among all the reference strains. Morphological differences are apparent between *V. anglica* 1589/11 and *Vannella* sp. F-49 (Nerad & Sawyer, pers. comm.), which indicates that although these isolates had identical riboprints they do not belong to the same species. Intraspecific variation was detected between the two *V. anglica* strains, although there were no polymorphisms between the two *P. plurinucleolus* strains.

Detection of restriction polymorphisms between closely related organisms is very much dependent on the sequence variation. Sequence analysis has demonstrated that there are no sequence differences among some species of *Tetrahymena*, and there is little variation within the mammalian blood fluke genus *Schistosoma* and the yeast genus *Zygosaccharomyces* (Sogin *et al.*, 1986c; Johnston *et al.*, 1993; James *et al.*, 1994). Restriction analysis of the *ssrRNA* gene from these genera will not

detect, or is very unlikely to detect, any interspecific differences. Conversely, restriction analysis can identify intraspecific variation between strains of the parasitic amoeba *Entamoeba histolytica* (Clark & Diamond, 1991a, 1991b), and strains of the amoeba-flagellate *Naegleria fowleri* (De Jonckheere, 1994a).

The detection of sequence variation also depends on the enzymes used. Within the schizopyrenid genus *Vahlkampfia*, only three restriction enzymes, *AluI*, *AvaII* and *BglIII*, out of a total of ten which digested the ssrRNA, were able to distinguish between all seven species (Brown & De Jonckheere, 1994). Within the *Naegleria* genus, there appeared to be less sequence variation, with only three enzymes, *AluI*, *HinfI* and *BanI*, out of a total of ten able to detect any polymorphisms among 14 strains from eleven species (De Jonckheere, 1994b). In both these cases *AluI* was an effective restriction enzyme, however, it only detected a single polymorphism within eight *Trypanosoma* species. In this case, the most effective enzymes for discriminating between species were *HaeIII*, *RsaI*, *Sau3AI* and *SerFI* (Clark *et al.*, 1995).

Considering that two enzymes could discriminate between all the named *Vannella* and *Platyamoeba* species, there appears to be significant sequence variation among these amoebae. In combination with morphological observations, riboprinting with as few as three or four enzymes may be an extremely useful tool for the unambiguous identification of amoebae isolated from the field.

### Identification of field isolates

Morphological observations of the floating and locomotive forms demonstrates that VP3 superficially resembles *P. bursella* and several marine *Vannella* isolates. The floating form pseudopodia of VP3 are rarely longer than the diameter of the central mass; considerably smaller than pseudopodia of *V. devonica* and *V. septrentionalis* (Page, 1979a, 1980a). The locomotive form of VP3 is larger than *V. sensilis* (Bovee & Sawyer, 1979), the average length: breadth ratio is smaller than *V. anglica* (Page, 1980a), and VP3 does not possess surface wrinkles which are apparent on *P. bursella* (Page, 1974a). The morphological observations of VP3 are most similar to *V. arabica*, although the floating form pseudopodia are less common and the locomotive form is marginally larger in *V. arabica* (Page, 1980a). Although the *HinfI* riboprints of VP3 and *V. arabica* are similar, the restriction data suggests that VP3 is most similar to *P. australis* and *V. anglica* 1589/8. However, the length of the

floating form pseudopodia, the greatest dimension and the length: breadth ratio of *P. australis* are considerably greater than VP3 which suggests that it does not belong to this species. There were considerable similarities between both the floating and locomotive forms VP3 and *V. anglica*, which suggests that VP3 is a strain of *V. anglica*. The only apparent differences were in the average length: breadth ratio, and the posterior edge of the locomotive form of *V. anglica* was more regular. However, considering that the riboprints of VP3 and *V. anglica* 1589/11 were quite different for both restriction enzymes, the identity of VP3 remains unclear. Together the morphological and molecular evidence suggests that VP3 is an undescribed member of the *Vannella* genus.

Species E is a reasonably small amoeba, and very few *Vannella* or *Platyamoeba* species of this size have been described. It was previously suggested that Species E may be a strain of *V. aberdonica* (Butler, 1994), however, although the shape of Species E floating form is irregular, it is not twisted or comma shaped like *V. aberdonica* (Page, 1980a), and small pseudopodia are present. Moreover, the locomotive form of Species E is quite unusual as it may adopt a tongue shape. Only two species have been described with a tongue or linguiform shape; a marine species, *P. leei*, and a freshwater isolate *P. stenopodia* (Page, 1969; Sawyer 1975b), but both these isolates are much larger than Species E. These observations suggested that Species E has not been previously described, but similarities indicated it is likely to be a *Platyamoeba* species. The riboprinting data demonstrated that Species E was substantially different from any of the *Vannella* and *Platyamoeba* reference species, and that this isolate may not be a member of either of these genera. However, comparisons were not possible with other small marine *Platyamoeba* species such as *P. langae*, *P. douvresi* and *P. muchelanoi* or the linguiform *P. leei* species which are not in culture.

Isolate SIA can be distinguished from most fan-shaped amoebae. The short floating form pseudopodia distinguish SIA from all freshwater *Vannella* isolates which have much longer pseudopodia (Page, 1988). The locomotive form is not linguiform, which distinguishes it from *P. stenopodia*, and the presence of small cysts demonstrate that SIA is not *P. schaefferi* or *Pessonnella marginata* which have large cysts or no cysts, respectively (Pussard, 1973; Singh & Hanumaiah, 1979). In all respects, the morphology of SIA and *Platyamoeba placida* are indistinguishable which suggest that SIA is a strain of this species (Page, 1969). Unfortunately, an insufficient

amount of the *ssrRNA* amplicon was available to confirm the identity of SIA by riboprinting.

### Phylogenetic analysis

The riboprinting data was subjected to a parsimony analysis to produce an unrooted most parsimonious phylogenetic tree. The use of two restriction enzymes was not expected to produce an accurate phylogenetic tree, but it was hoped that it would provide an indication of the relationships between some of the isolates, and that it would demonstrate how the technique could be used to interpret riboprinting data for phylogenetic purposes. However, the low bootstrap values indicated that the use of two restriction enzymes was clearly insufficient to resolve any phylogenetic relationships.

Restriction analysis using between ten and twelve restriction enzymes which recognise a four base pair sequence provide a reasonably accurate phylogenetic analysis of closely related species (Clark & Diamond, 1991a; Clark, 1992; Brown & De Jonckheere, 1994; Clarke *et al.*, 1995). With the use of more enzymes it is also possible to improve the accuracy of a phylogeny by mapping restriction sites, by using sites as independent characters rather than co-migrating restriction fragments (Clark & Diamond, 1991a; Clark, 1992). Co-migrating fragments are not necessarily homologous, and assumptions of independence among characters are violated. For example, if a new site evolved between two existing sites, the longer fragment would disappear and two shorter ones would take its place. Therefore, even though two species have two of the three restriction sites, they have no common fragments (Swofford *et al.*, 1996). This is a potential source of error, although it is argued that if enough restriction enzymes are examined phylogenies are not significantly affected (Bremer, 1991).

To provide a reasonably accurate phylogeny of the fan-shaped amoebae, eight or more restriction enzymes are therefore required, and the use of restriction sites rather than co-migrating fragments would be favoured. However, as substantial quantities of amplified *ssrRNA* genes were not available, an alternative approach was used to examine the phylogeny of these amoebae. From the analysis of the primary and secondary *ssrRNA* sequence from *V. anglica* (Chapter 3) and other eukaryotic *ssrRNAs* (Gutell, 1994; Maidak *et al.*, 1996; Van de Peer *et al.*, 1994;), a variable region of the gene was selected for partial DNA sequencing. The variable

region was amplified, cloned and sequenced, and the sequence data were used to determine the phylogeny of the fan-shaped amoebae (Chapter 5).

The usefulness of riboprinting for the identification and examination of phylogenies among closely related organisms is very much dependent on the genetic diversity of the *ssrRNA* gene. Riboprinting of the *Vannella* and *Platyamoeba* isolates has demonstrated that there is a significant amount of genetic variation between these morphologically similar isolates. Riboprinting with just two enzymes has provided a valuable insight into the identity of VP3 and Species E. Considering the lack of rigid diagnostic features available at the morphological level, riboprinting provides a valuable additional tool for the identification of these amoebae. Further improvements in the PCR methodology would allow rapid identification of large numbers of cultured amoebae or organisms isolated from the field. It could be a promising approach in biogeographical studies where the unambiguous identity of morphologically similar isolates from different sites is difficult to determine using conventional approaches. The use of more restriction enzymes would also allow the phylogenetic relationships between these amoebae to be determined, so the evolution of these organisms and their characters could be examined.



## **CHAPTER 3**

### **Cloning, sequencing and secondary structure of the** ***Vannella anglica* ssrRNA**

## Introduction

### **Primary sequence of ssrRNA**

Comparative sequence analysis has shown that the ssrRNA gene sequence is chimeric in nature, composed of highly conserved regions, semi-conserved regions and variable regions. The highly conserved regions are present universally, and are essentially invariant. Sequence elements in these regions are involved with various aspects of the structure and function of the rRNA molecule. The semi-conserved regions are also responsible for maintaining the secondary structure of the molecule. Although there is sequence variation in these regions, compensatory base changes ensure that the secondary structure is very similar (Gutell *et al.*, 1986). The sequence variation in the variable regions is significantly greater, but, homologous structural elements are still evident. Extensive insertions, or expansions segments, are present within the variable regions of some taxa which account for the large differences in the size of ssrRNAs (Gunderson & Sogin, 1986; Hinkle *et al.*, 1994).

### **Secondary structure of ssrRNA**

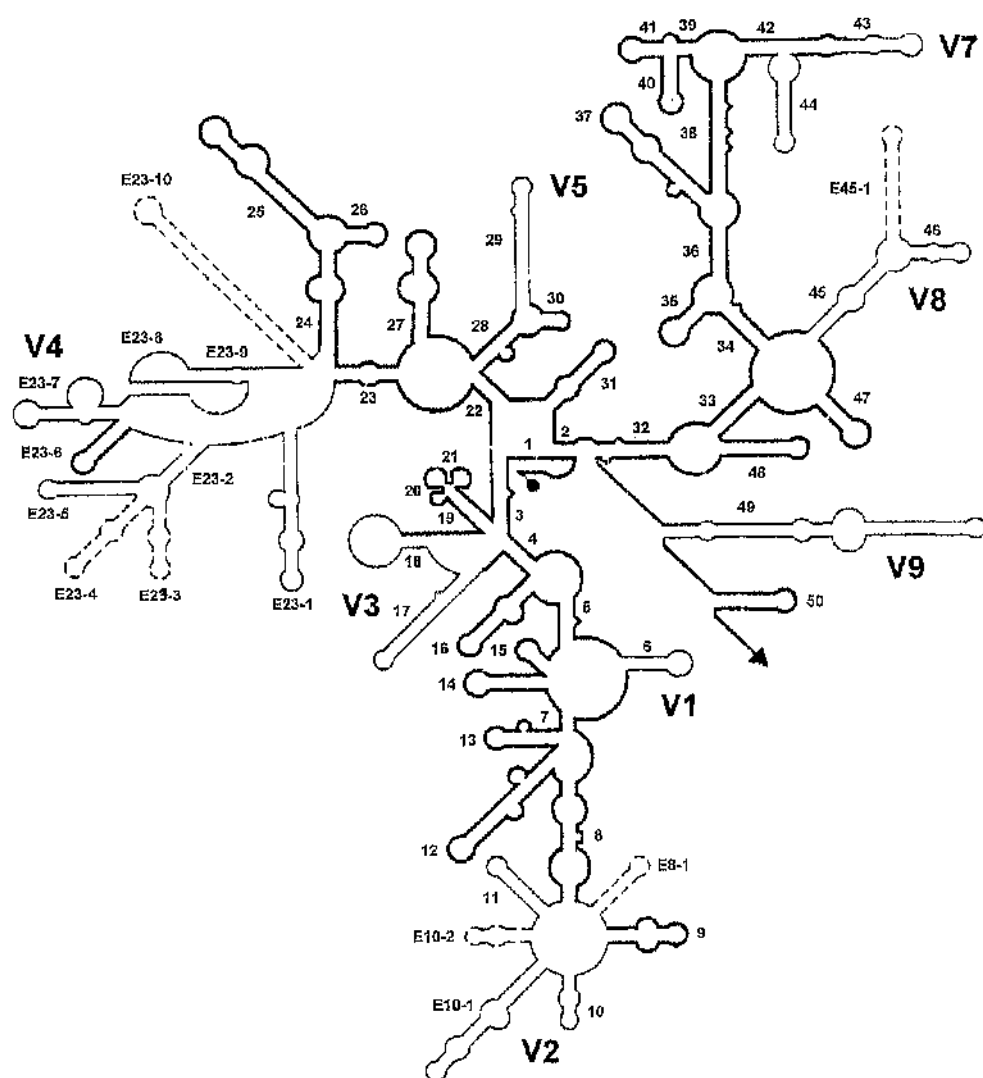
Non-covalent interactions between bases in RNA result in the formation of secondary and higher-order structure (Noller & Woese, 1981; Gutell *et al.*, 1986). Most of the secondary structure interactions are between canonical base pairs i.e. between guanine and cytosine (G:C), and between adenine and uracil (A:U), however non-canonical base-pairing between guanine and uracil (G:U) is also common. Comparative sequence analysis suggests that other non-canonical interactions occur infrequently, although guanine-adenine (G:A) interactions are the most common class (Gutell *et al.*, 1994). Consecutive base-pairings between RNA sequence elements produce helices, most of which are completed with a terminal or hairpin loop of between three and six bases. Helices are often interrupted by internal loops, unilateral bulges or non-complementary bases. Helices can be confirmed using comparative sequence analysis. The presence of two or more independent pairs of compensating base changes is usually sufficient to describe a helix as "proven" (Gutell *et al.*, 1985).

Comparative studies of RNA gene sequences led to the discovery of the characteristic "clover-leaf" configuration of tRNA (Holley *et al.*, 1965; Zachau *et al.*, 1966), and examination of co-variation using a more extensive set of tRNA

sequences in later studies identified a few higher order structural elements (Levitt, 1969). However, a majority of the secondary and tertiary interactions in tRNA structure were determined by X-ray crystallography (Kim, 1979). Nevertheless, comparative analysis with the comprehensive tRNA sequence database now available, allows all the secondary structure interactions and some tertiary interactions to be detected (Gutell *et al.*, 1992). Examination of rRNAs by X-ray crystallography has proven considerably more difficult than tRNAs because the ribosomal subunits are relatively unstable, crystals are difficult to grow and the diffraction resolution is poor (Yonath *et al.*, 1990). In the absence of X-ray data, comparative sequence analysis of rRNA sequences has been instrumental for inferring the secondary structures of 5S, 16S and 23S rRNA molecules (Fox & Woese, 1975; Noller & Woese, 1981; Noller *et al.*, 1981; Woese *et al.*, 1983; Gutell *et al.*, 1985).

The generalised secondary structure model for nuclear eukaryotic ssrRNAs is shown in Figure 3.1. Regions of high sequence conservation are depicted as bold lines, while thin lines represent regions of variable sequence. Previous sequence analyses identified seven ssrRNA variable regions (Sogin & Gunderson, 1987), however, a more recent scheme, which was adopted here, recognises nine separate variable regions known as V1 to V9 (De Rijk *et al.*, 1992). The V6 region is only present in prokaryotic ssrRNAs. This region, corresponding to helix 37, is not considered variable in eukaryotic ssrRNAs. There are 50 homologous helices which occur in prokaryotes and eukaryotes, almost without exception. These otherwise universal helices are numbered from the 5' terminus as they are encountered (Neefs *et al.*, 1993). Previous models only recognised 48 helices (De Rijk *et al.*, 1992), however, a pseudoknot formation recognised within helix 19 produced two additional small helices (Woese & Gutell, 1989). These small helices are numbered 20 and 21, and the number of all the succeeding helices have now been increased by two (Neefs *et al.*, 1993).

The Microsporidia taxa provide the most unusual eukaryotic ssrRNA secondary structures. Organisms belonging to this group lack several universal helices, and most if not all the helices which are specific to eukaryotes. *Viarimorpha necatrix* lacks helices 10, 11, 43 and 46, and helices 11, 18, 43 and 46 are absent from *Encephalitozoon cuniculi*. With respect to the eukaryotic specific helices, *E. cuniculi* possesses only helices E23-8 and E23-9, and *V. necatrix* possesses none (Vossbrinck



**Figure 3.1** Secondary structure model for eukaryotic ssrRNAs (De Rijk *et al.*, 1992). The 5' terminus is symbolised by a dot, and the 3' terminus by an arrowhead. The helices are numbered in order of occurrence from the 5' terminus. Conserved regions are shown in bold, and variable regions labelled V1 to V9 are drawn in thin lines. Helices prefixed with an 'E' are eukaryotic specific and helices drawn in broken lines are only present in some eukaryotic taxa.

*et al.*, 1987; Hartskeerl *et al.*, 1993). Almost all other eukaryotic taxa have at least helices E10-1, E23-1, E23-2, E23-5, E23-6, E23-7, E23-8 and E23-9. Helix E-8 is present in the ciliate *Euplotes aediculatus*, the euglenoid *Euglena gracilis*, the plasmodial slime mold *Physarum polycephalum* and most kinetoplasts. Helix E10-2 has only been found in the amoeba-flagellate *Naegleria gruberi*. Helices E23-3 and E23-4 have been found in Platyhelminths, insects and the euglenoid *E. gracilis*. Helix E23-4 is also present in *N. gruberi* and the naked amoeba *Acanthamoeba castellanii*, but not *Hartmannella vermiformis*. Helix E23-10, and perhaps other helices between E23-9 and 24, are present in *E. gracilis*, and most kinetoplasts. Moreover, helix E43-1 is also present in the amoeba-flagellate *N. gruberi*, the naked amoeba *A. castellanii*, the apicomplexa genus *Plasmodium*, but few other taxa (Neefs *et al.*, 1993).

### Higher-order Interactions

In addition to the secondary structure elements which essentially include only regular helical elements composed of canonical and G:U base-pairs, there are other structural elements such as pseudoknots, lone-pairs and non-canonical pairs, which are collectively known as tertiary or higher-order interactions. Pseudoknots occur when hairpin loops interact with bases which are external to the hairpin. Three examples are apparent in the generalised ssrRNA model shown in Figure 3.1. The first pseudoknot occurs at the 5' terminus, between the hairpin loop of helix 1 and the region between helices 31 and 32, to form helix 2. The second pseudoknot, mentioned previously, is a more complicated interaction within helix 19 resulting in the formation of helices 20 and 21 (Woese & Gutell, 1989; Powers & Noller, 1991). Due to the variable size and sequence composition, the secondary structure of the V4 region has been difficult to determine (Nickrent & Sargent, 1991). However, examination of compensatory changes uncovered a third pseudoknot between helices E23-8 and E23-9 (Neefs & De Wachter, 1990), which has significantly aided the elucidation of the structure of this region. Finally, a fourth pseudoknot, not shown in Figure 3.1, has been proposed for an interaction involving two bases from the hairpin loop of helix 30 and two bases from the single-stranded region between helices 22 and 23 (Gutell *et al.*, 1986).

Comparative sequence analysis has revealed several higher-order interactions involving lone pairs (Gutell *et al.*, 1985, Gutell, 1993). However, interactions of this type are likely to go unnoticed if they involve bases whose composition seldom

varies or if they involve non-canonical base-pairing. Non-canonical base-pairing may also be important immediately beyond the 5' and 3' termini of established helices. Interactions of this type have been inferred to extend helices 47 and 49 (Gutell *et al.*, 1994).

### **Structure-function relationship**

Numerous studies have been conducted on the structure-function relationships of ssrRNA in prokaryotes with a view to determining the role of the small ribosomal subunit in translation (Hill *et al.*, 1990). Examination of naturally occurring mutations, which provided resistance to antibiotics which otherwise block specific steps in the translation process, provided the first direct evidence that rRNAs, rather than the ribosomal proteins, play the fundamental role in translation (De Stasio *et al.*, 1987). Site-directed mutagenesis in combination with functional analysis of mutant rRNA, has allowed individual nucleotides to be identified which are directly involved with the translation process (Morgan *et al.*, 1987; Moazed & Noller, 1987; De Stasio *et al.*, 1989). Together studies on prokaryotes, particularly *E. coli*, have identified sites involved with initially binding mRNA, subunit association, tRNA binding, decoding of mRNA, and elongation and termination of polypeptide synthesis (Raué *et al.*, 1990). Comparisons with eukaryotic ssrRNA sequences indicate that homologous regions are present in eukaryotes, although there is currently little direct evidence. However, systems for the direct examination of eukaryotic rRNA function are being developed for *Tetrahymena* and yeast (Sweeney *et al.*, 1993; Venema *et al.*, 1995).

### **Ribosomal RNA sequencing strategies**

Over the years, a variety of strategies have been used to sequence ribosomal RNA (Hillis *et al.*, 1996). Essentially, rRNA genes can be cloned and sequenced, sequenced directly from RNA or sequenced directly from a PCR amplification of the gene. In early studies, generally before the development of a ribosomal RNA database and the advent of the polymerase chain reaction, ribosomal RNA genes were isolated from genomic DNA libraries, sub-cloned into plasmid or M13 bacteriophage vectors and sequenced (McCarroll *et al.*, 1983; Gunderson *et al.*, 1987; Eschbach *et al.*, 1991; Kwon *et al.*, 1991). As sequence data became available, primers were designed for direct sequencing of rRNA using reverse transcriptase. This method was popular for phylogenetic reconstructions and for the rapid identification of micro-organisms (Lane *et al.*, 1985; Böttger, 1989). Since the

development of the polymerase chain reaction, hundreds of rRNA genes have been amplified, cloned into single-stranded phage (Medlin *et al.*, 1988; Schlegel *et al.*, 1991; Hinkle *et al.*, 1994; Leipe *et al.*, 1994) or plasmid vectors (Hendriks *et al.*, 1990; Fong *et al.*, 1993; McFadden *et al.*, 1994b), or sequenced directly (Embley, 1991; Tan & Druehl, 1993; Weekers *et al.*, 1994; Zhu *et al.*, 1994).

### **Cloning versus direct sequencing**

The speed of direct sequencing is perhaps the greatest advantage over approaches which entail an intermediate cloning step. Furthermore, to obtain an unambiguous sequence of a cloned amplified gene requires the analysis of multiple clones as DNA polymerases may have introduced errors. Direct sequencing of amplified genes reduces problems associated with errors in all but the earliest rounds of amplification. However, there are disadvantages for direct sequencing strategies such as the lack of 'hard' copies of the amplicon for sequence verification or other purposes, and the requirement for sequence information in flanking regions. Moreover, in direct RNA sequencing, only transcribed regions are accessible, no complementary strand is available for verification of the sequence, and sequencing may terminate prematurely due to the formation of secondary RNA structures. Direct sequencing of PCR products is also limited to relatively short regions, and the length of the amplification product must be homogeneous. Most of the disadvantages of directly sequencing amplicons, can be overcome by combining PCR with cloning techniques. Cloning amplified products is considerably quicker than cloning from genomic DNA, especially when the cloning methods have been specifically designed for this purpose (Hillis *et al.*, 1996).

### **Cloning PCR-amplified genes**

Initial attempts to directly clone amplicons into blunt-ended vectors were inefficient, due to template-independent terminal transferase activity of *Taq* DNA polymerase (Clark, 1988; Mole *et al.*, 1989). *Taq* DNA polymerase adds a single nucleotide, almost exclusively an deoxyadenylate residue to the 3' hydroxyl termini of double-stranded DNA, due to the strong preference of the polymerase for dATP (Clark, 1988). In the light of this finding, a variety of techniques have been developed to improve the cloning efficiency of amplification products.

One technique removes the single deoxyadenylate extension from the 3' termini of the PCR product, prior to ligation into a blunt-ended vector (Hemsley *et al.*, 1989).

This can be achieved by the 3' to 5' exonuclease activity of the Klenow fragment of DNA polymerase I. Additional treatment of the vector with alkaline phosphatase is often required to reduce religation of the vector and the subsequent high frequency of non-recombinant clones.

Endonuclease restriction sites may be incorporated into each of the oligonucleotide amplification primers to assist cloning (Scharf *et al.*, 1986). Digestion of the amplicon, with the restriction enzyme(s), produces a double-stranded DNA fragment with cohesive ends which can be ligated to a vector cut with the same enzyme(s). The method, is clearly dependent upon the absence of the selected restriction sites within the amplified fragment, but other difficulties may also be encountered. Some endonucleases are inefficient at digesting DNA if the recognition site is located near the termini of the DNA fragment (Kaufman & Evans, 1990). This problem may be alleviated by increasing the length of the primer, however, this may impair PCR efficiency. Alternatively, phosphorylated PCR products may be ligated into concatamers with DNA ligase prior to endonuclease restriction (Jung *et al.*, 1990; Kaufman & Evans, 1990).

Alternative strategies have been developed which do not require modification of the amplification product or large primer extensions. These methods allow direct cloning of amplicons, using vectors with single complementary 3' deoxythymidylate termini, which provide a cohesive substrate for the efficient ligation with 3' deoxyadenylate amplicons. These T-extended vectors may be produced by the addition of a 3' deoxythymidylate to blunt-ended vectors using dd'TTP in the presence of termini transferase (Holton & Graham, 1990), or *Taq* DNA polymerase (Marchuk *et al.*, 1990). Plasmid vectors have been developed which produce single 3' deoxythymidylate nucleotide extensions when digested with *Hph*I or *Xcm*I restriction enzymes (Mead *et al.*, 1991; Kovalic *et al.*, 1991). These vectors provide cohesive termini for the direct ligation of unmodified 3'-deoxyadenylate PCR products. Amplicons up to 2.4 kbp have been cloned using this experimental procedure at an efficiency 50 times greater than blunt-end cloning (Mead *et al.*, 1991).

These T-extended plasmids, otherwise known as TA-cloning plasmids, have a variety of advantages over other PCR product cloning methods. With these plasmids, amplicons may be cloned directly without removal of overhanging 3' deoxyadenylate



residues, or the addition of deoxythymidylate to the 3' termini of linearised vector. Extended primers do not have to be designed and no prior knowledge of the amplified sequence is required. However, with TA-cloning there may be a high incidence of background transformation. To circumvent this problem, recombinant clones are screened for using  $\alpha$ -complementation of the *lacZ* gene. On media containing the chromogenic substrate X-Gal, the presence of a DNA insert disrupts the *lacZ* gene resulting in white colonies, whereas non-recombinant clones produce blue colonies.

### **DNA sequencing**

In 1977, two quite different methods were published for the sequencing of DNA molecules (Maxam & Gilbert, 1977; Sanger *et al.*, 1977). The Maxam-Gilbert method uses chemicals to bring about base-specific modification and cleavage. A DNA strand is radioactively labelled at one end and divided into four samples. To each sample chemicals are added under certain conditions to modify a small proportion of one or two of the bases. Piperidine is then used to cleave DNA strand at the modified base (Maxam & Gilbert, 1977, 1980). Random modification and cleavage of the large population of DNA fragments ensures that DNA is cleaved at each nucleotide position. The radioactively labelled DNA fragments are separated by electrophoresis by size on a polyacrylamide gel, and visualized by exposing the dried gel to X-ray film to produce an autoradiograph from which the sequence of the DNA fragment can be read directly. This method is no longer used extensively, but it may be useful for determining the DNA sequence of RNA genes when the formation of secondary structure is problematic.

The Sanger method, also known as the chain-terminating method, uses dideoxynucleotide analogues in primer-directed DNA polymerization to produce discrete DNA fragments (Sanger *et al.*, 1977). Initially, oligonucleotide primers are annealed to a complementary region of single-stranded template DNA. The DNA is divided into four samples and four deoxynucleotides (dATP, dTTP, dGTP and dCTP), one of which is radioactively labelled, are added to sample. In addition, one of four dideoxynucleotides (either ddATP, ddTTP, ddGTP or ddCTP) is added to each of the sub-samples. The DNA is extended from the free 3'-hydroxyl group of the primer sequence by DNA polymerase using the target DNA as a template. At some stage a dideoxynucleotide will become incorporated into each extending DNA fragment and the polymerization will terminate because dideoxynucleotides lack a free 3'-

hydroxyl group, which is essential for DNA elongation. The population of DNA fragments for each of the four radioactively labelled samples may then be separated by electrophoresis and visualized by autoradiography.

The Sanger chain termination method is considered better than the Maxam-Gilbert method, and derivatives of the Sanger method, such as cycle sequencing and automated sequencing, are an active area of development. Unlike the Maxam-Gilbert method, which requires knowledge of the restriction map so the target DNA can be digested into manageable fragments for sequencing, no prior knowledge of the target sequence is necessarily required for Sanger sequencing. Sequences complementary to the vector or amplification primers can be used to initiate DNA synthesis, but perhaps the most attractive feature of Sanger sequencing is that more sequence can be read due to the better band resolution.

Cycle sequencing is based on the chain termination method of Sanger, but utilises a linear polymerase chain reaction to amplify labelled DNA that is complementary to the target DNA (Murray, 1989; Craxton, 1991). In thermal cycle sequencing, DNA synthesis is catalysed by thermostable DNA polymerase. Heat denaturation of double stranded template allows labelled primers access to a single strand and subsequent extension by the polymerase. Successive cycles of denaturation, annealing and synthesis result in a linear amplification of labelled product. The band resolution of this method has been improved by end-labelling the primer with a radioactive phosphate group using T4 polynucleotide kinase, using biotinylated primers for chemi-luminescent DNA sequencing (Creasey *et al.*, 1991) or end-labelling with fluorescent dyes for automated DNA sequencing (Tracey & Malcahy, 1991). Cycle sequencing is a very versatile method as template DNA can be single-stranded or double-stranded, from cloned sources or direct from PCR amplifications. Incorporation of labelled primer is very effective in providing clear resolution from relatively small amounts of DNA. Moreover, many sequencing artifacts are removed during the thermal cycling. However, this method is susceptible to problems characteristic of PCR, including the standardization of conditions across thermal cycling machines, the establishment of optimal conditions for different primer sets and the potential incorporation of further errors by thermostable DNA polymerases.

Automated sequencing usually uses a chain termination sequencing method in conjunction with fluorescently labelled DNA fragments which are detected by a laser

as they migrate through the polyacrylamide gel during electrophoresis. The process is described as automated since visual inspection of an autoradiograph is not required; instead the sequence is recorded directly into a computer. Automated sequencing saves significant amounts of time and alleviates clerical errors, but is expensive and requires professional maintenance.

### **Summary of chapter**

In this chapter, the PCR-amplified *ssrRNA* from *V. anglica* was cloned into a TA-cloning vector and the complete *ssrRNA* gene sequence was determined using the Sanger dideoxynucleotide sequencing method. Comparative analyses of other eukaryotic *ssrRNAs* was used to determine the secondary structure of the *V. anglica* *ssrRNA*. Interesting secondary structure features, predicted higher order interactions, and possible errors are discussed.

## **Materials and methods**

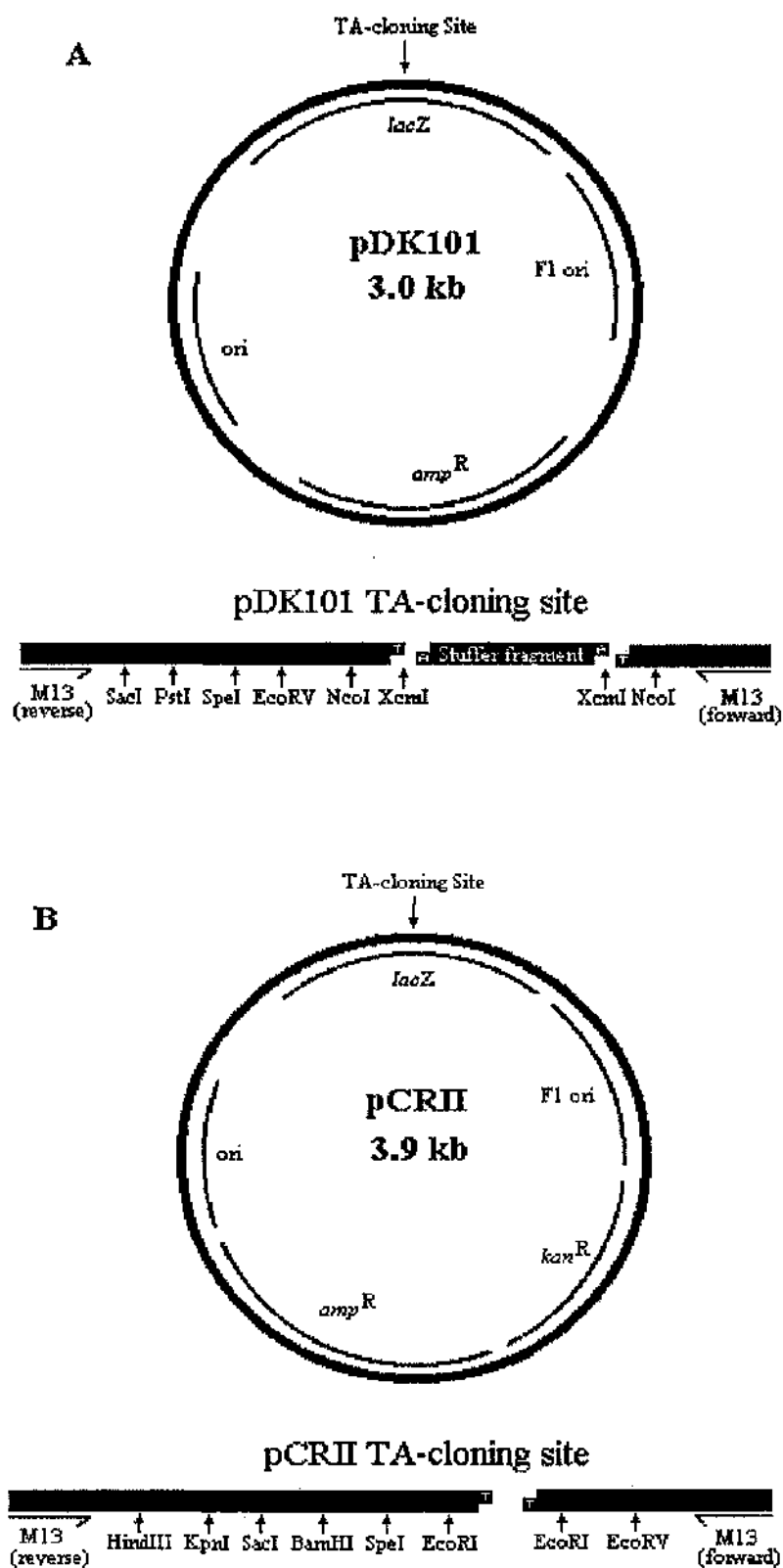
### **TA-cloning**

Attempts were made to clone *ssrRNA* genes using two TA-cloning systems. Initially, *ssrRNA* amplicons were cloned into plasmid pDK101, a 3 kbp plasmid which produces single 3' deoxythymidylate overhangs when digested with the restriction enzyme *XcmI*. This plasmid, however, was succeeded by the use of pCRII, a commercially prepared 3.9 kbp linearised plasmid, part of Invitrogen's TA-Cloning System.

Details of pDK101 and the pCRII TA-cloning plasmids are shown in Figure 3.2 A and B, respectively. Plasmid pDK101 was constructed from the pGEM5Zf(+) plasmid by incorporating a pair of *XcmI* restriction sites within the multiple cloning site (Kovalic *et al.*, 1991). When digested with *XcmI*, pDK101 produces a 15 bp stuffer fragment and linearised plasmid of approximately 3 kbp with incompatible single 3' deoxythymidylate overhangs, suitable for the direct cloning of genes amplified by *Taq* DNA polymerase. Invitrogen's prepared pCRII plasmid DNA is linear with single 3' deoxythymidylate nucleotide overhangs. Both plasmids encode antibiotic resistance genes for selection of transformed cells, and the TA-cloning sites are located within the amino-terminal of the *LacZ* gene for  $\alpha$ -complementation screening colonies with recombinant plasmids.

### **Preparation of pDK101 *XcmI* 3Kbp fragment**

Ten millilitres of SOC media (Appendix I) containing 100  $\mu$ g/ml of ampicillin was inoculated with a single colony of *Escherichia coli* strain CC118 pDK101 and incubated overnight at 37°C in a shaking water-bath. Plasmid pDK101 was isolated from 3 ml of culture using Promega's Wizard Mini-prep DNA purification systems. One microgram quantities of plasmid DNA were digested with 10 units of *XcmI* (New England Biolabs) for 2 hours at 37°C. The linear 3 kbp plasmid was separated from the stuffer fragment by electrophoresis using low melting point agarose. The pDK101 *XcmI* fragment was excised from the gel and purified using GeneClean II (BIO101).



**Figure 3.2** TA-cloning plasmids with cloning site details. (A) pDK101, and (B) pCRII.

### **Ligation of linear plasmid DNA and PCR products**

Approximately 50 ng of pDK101 *Xcm*I preparation and 50-250 ng of purified *ssr*RNA amplicon were ligated with 10 Weiss units of T4 DNA ligase (New England Biolabs) in 30  $\mu$ l of 1x ligase buffer, at 16°C for 16 hours or 25°C for 2 hours. For the pCRII TA-cloning plasmid, ligation reactions were performed as recommended by the manufacturer. Fifty nanograms of linear pCRII vector and 1  $\mu$ l of amplified *ssr*RNA gene were ligated with 4 Weiss units of T4 DNA ligase in 10  $\mu$ l of buffered solution at 12°C overnight.

### **Transformation of plasmids and selection of recombinant colonies**

A variety of competent cells were used to transform plasmids all of which allow for  $\alpha$ -complementation of recombinant plasmids. For pDK101, competent Library Efficiency DH5 $\alpha$  cells (Gibco BRL) were used and for pCRII, competent One Shot cells (Invitrogen's TA-Cloning Kit) and Max Efficiency DH5 $\alpha$ F'I<sup>q</sup> (Gibco BRL) were used. Transformations were carried out as recommended by the manufacturers. Essentially, competent cells were thawed on ice and 20-50  $\mu$ l of cells and aliquots of the ligation mixture were added, gently mixed with a pipette tip and left on ice for 30-45 min. The transformation mixtures were transferred to a 42°C block for 30 sec to heat-shock the cells and placed on to ice for 2 min. The mixtures were made up to 500  $\mu$ l with liquid SOC media (Appendix I) and incubated for 1 hour at 37°C in a shaking water bath to allow expression of the ampicillin resistance gene. Aliquots of the cultures were plated onto nutrient agar plates containing 100  $\mu$ g/ml ampicillin overlaid with 40  $\mu$ l of 20 mg/ml X-Gal (Sambrook *et al.*, 1989). For transformations involving Max Efficiency DH5 $\alpha$ F'I<sup>q</sup> cells in addition to X-Gal, 4  $\mu$ l of 200 mg/ml IPTG was also spread over the plate to allow expression from the *lac* promoter. Plates were incubated at 37°C overnight. White and blue colonies were enumerated and transformation frequencies were calculated. A proportion of the transformed colonies were transferred to master-plates to confirm the colony colour phenotype.

### **Analysis of colonies with suspected recombinant plasmids**

Plasmid DNA from colonies with potentially recombinant plasmids was examined by restriction analysis. Plasmid DNA was isolated from overnight cultures using Wizard Mini-prep Systems (Promega). The plasmid DNA was digested at the cloning site to linearise the plasmid or to excise the DNA insert, and the size of the DNA fragments were determined by agarose gel electrophoresis. Potentially recombinant

pDK101 plasmids were digested with *EcoRV*, and pCRII plasmids were digested with *EcoRI*. Suspected recombinant clones from the pCRII cloning were labelled with the prefix pCRV01.

### Sub-cloning *V. anglica* *ssrRNA* gene

To aid DNA sequencing, the TA-cloned *V. anglica* *ssrRNA* gene was sub-cloned. Initially, the plasmid containing the *ssrRNA* gene was digested with a series of restriction enzymes to identify enzymes which divided the gene into more manageable fragments. In combination, *EcoRI* and *Bam*HI restriction enzymes divided the gene into three fragments of approximately 1.0, 0.7 and 0.2 kbp. The *ssrRNA* gene fragments and recipient plasmid DNA fragments were purified using agarose gel electrophoresis and Geneclean II. The 1.0 kbp and 0.7 kbp *EcoRI* *Bam*HI fragments were cloned into 2.7 kbp plasmid pUC18 which was cut similarly, and white colonies with suspected recombinant plasmids were labelled with the prefixes pUCV11 and pUCV12, respectively. The 0.2 kbp *EcoRI* fragment was cloned into pCRII *EcoRI* and the white colonies were labelled with the prefix pCRV03 (See Figure 3.3). Ligations were carried out using approximately 50 ng of linear plasmid DNA, 100-200 ng of *ssrRNA* gene fragment and 10 Weiss units of DNA ligase, at 25°C for 1 hour. Transformations and selection of recombinant colonies were performed as described previously.

### DNA sequencing

Chain terminating DNA sequencing was performed using the Sequenase Version 2.0 DNA sequencing kit and <sup>35</sup>S-labelled dATP (Amersham Life Science). This sequencing kit permits the use of single-stranded or double-stranded templates, and includes reagents for reading sequence close to the primer, for extending sequence further from the primer and eliminating sequence compressions caused by the formation of secondary structures. The sub-cloned *V. anglica* *ssrRNA* gene fragments was sequenced with double-stranded plasmid DNA using M13 primers flanking the *ssrRNA* gene inserts, and internal primers complementary to conserved regions of *ssrRNAs*. The internal primers were designed using an alignment of *ssrRNA* sequences from the amoebae *Phreatamoeba balamuthi* (Hinkle *et al.*, 1994), *Acanthamoeba castellanii* (Gunderson & Sogin, 1986), *Hartmannella vermiformis* (Gunderson *et al.*, 1994), the apicomplexa *Sarcocystis muris* (Gajadhar *et al.*, 1991), the dinoflagellate *Symbiodinium corculorum* (McNally *et al.*, 1993), the cryptomonad *Cryptomonas phi* (Douglas *et al.*, 1991), the red alga *Gracilaria lemaneiformis*

(Bhattacharya *et al.*, 1990) and the green alga *Klebsormidium flaccidum* (Wilcox *et al.*, 1993). The *ssrRNA* gene sequences were obtained from the GenEMBL database using FETCH and aligned using PILEUP and PRETTY programs which are available from the University of Wisconsin Genetics Computer Group (GCG) sequence analysis package (Devereux *et al.*, 1984). The sequencing strategy and the list of primers used are shown in Figure 3.3.

### **Preparation of template DNA**

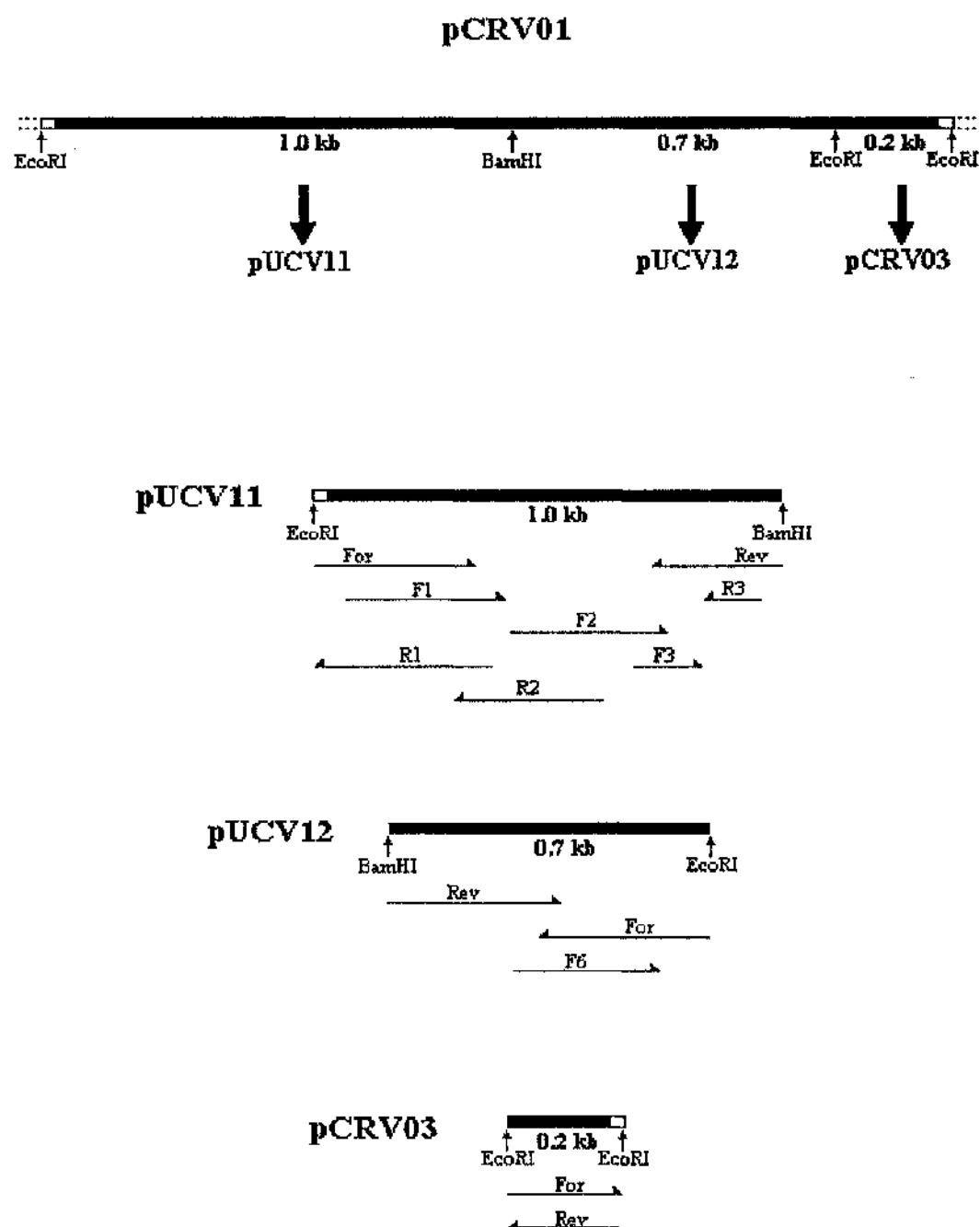
Plasmid DNA containing the *ssrRNA* gene fragments was isolated using Wizard Mini-prep systems. The double-stranded DNA was denatured by adding 0.1 volumes of 2 M sodium hydroxide and 2 mM EDTA, and incubating for 30 min at 37°C. The mixture was neutralized with 3 M sodium acetate (pH 4.8) and the DNA was precipitated with two volumes of ethanol for 30 min at -70°C. The DNA was pelleted by centrifugation, washed four times with 70% ethanol and stored dry at -20°C for no more than one day before use.

### **Sequencing reactions**

Sequencing reactions were performed as recommended by the manufacturer. The DNA sequencing reaction involved two steps. In the first step, the primer was annealed to the template DNA and with limited amounts of dNTPs, <sup>35</sup>S-labelled dATP was incorporated into DNA chains which vary in length from several nucleotides to hundreds of nucleotides. In the second step, the concentration of deoxynucleotides was increased to extend the DNA chains and low concentrations of dideoxynucleotides were introduced to terminate the extended DNA fragments.

For each sequencing reaction one freshly prepared mini-prep of denatured, dried plasmid DNA was resuspended with 1 µl of primer DNA (1 pmol), 2 µl of reaction buffer and 10 µl of sterile water. The mixture was immediately heated to 65°C for 2 min and cooled to 30°C over 15-30 min to allow the primer to anneal to the template DNA. The mixture was cooled on ice before 1 µl of 0.1 M DTT, 2.7 µl of labelling mix, 0.5 µl of <sup>35</sup>S-dATP and 2 µl of diluted Sequenase enzyme were added. The labelling mixture was mixed, avoiding the formation of air bubbles, and incubated at room temperature for 2-5 min. Five microlitres of the labelled mixture was transferred to four labelled tubes containing 4 µl of one of the four termination mixtures, either ddATP, ddCTP, ddGTP or ddTTP. All four termination reactions were mixed thoroughly and incubated at 37°C for 5 min. The termination reactions





#### Flanking primers

For (M13 -40 forward primer)	5'	GTTTTCCCAGTCACGA	3'
Rev (M13 reverse primer)	5'	CAGGAAACAGCTATGAC	3'

#### Internal primers

F1	5'	CTGGTTGATCCTGCCAG	3'	R1	5'	TCTCCGGAATCGAACCC	3'
F2	5'	AGGGTTCGATTCCGGAG	3'	R2	5'	GGAATTACCGCGGCTGG	3'
F3	5'	CGGTAATTCCAGCTCCAA	3'	R3	5'	CCAAGAATTTACCTCTG	3'
F6	5'	TGGTGCATGGCCGTTT	3'				

**Figure 3.3** Sub-cloning and sequencing strategy for the TA-cloned *V. anglica* ssrRNA gene.

were stopped by the addition of 4  $\mu$ l of Stop Solution, mixed thoroughly and stored frozen until the sequencing gel was ready to be loaded.

### **Gel casting, running conditions and autoradiography**

Standard sized glass sequencing plates were initially washed with warm water and Ajax, followed by two subsequent washing steps with ethanol, one with Gel Slick (AT Biochem) and a final wash with ethanol. The plates were separated by 0.4 mm thick plastic spacers and sealed together using heat resistant tape. To 70 ml of premixed 6% w/v acrylamide, 0.3% w/v bis-acrylamide with 7 M urea (Easigel, Scotlab), 140  $\mu$ l of 25% ammonium persulphate and 70  $\mu$ l of TEMED were added and mixed. Without delay, the acrylamide solution was carefully poured between the plates to avoid air bubbles and left to polymerize for at least 90 min.

Prior to loading, the four termination reactions were pre-heated to 75°C for 10 min, then 2  $\mu$ l from each was loaded in consecutive lanes. The gels were run on a manual sequencing apparatus for 2.5 and 4.5 hours in 1x TBE buffer at a constant current of 60 amps. The gel was allowed to cool, transferred to 3 MM chromatography paper, overlaid with Saran Wrap and dried under vacuum at 80°C for 4 hours on a gel drier. The DNA sequence was visualised by exposing against X-ray film for 1-7 days.

### **Primary sequence analysis and secondary structure determination**

The DNA sequences were read, and compiled on computer using the multi sequence editor ESEE (Cabot & Beckenbach, 1989). The size of the gene, the number of guanine and cytosine nucleotides, as a proportion of the total number of nucleotides (G+C content), and the position of previously observed restriction sites were all noted. The primary ssrRNA of *V. anglica* was initially aligned with other eukaryotic ssrRNA sequences using the highly conserved sequence elements. Helices were identified and determined by comparisons of the sequence alignment with the compilations of ssrRNA secondary structure predictions (De Rijk *et al.*, 1992; Gutell, 1993, 1994; Neefs *et al.*, 1993; Van de Peer *et al.*, 1994, 1996). To aid the determination of helices where sequence conservation was poor, the RNA secondary structure program FOLDRNA was also employed (Zuker, 1989). Potential higher-order interactions, such as pseudoknots, non-canonical base-pairs and extended helices were also examined. Sequence errors which may have been incorporated by DNA polymerase during amplification or DNA sequencing were also noted.

## Results

### **TA-cloning using pDK101**

The TA-cloning results shown in Table 3.1 are typical of those obtained with pDK101. Plasmid pUC18 yielded transformed DH5 $\alpha$  cells at a frequency of approximately  $5 \times 10^6$  transformants per microgram of plasmid DNA, at least ten-fold higher than the TA-cloning plasmid pDK101. Transformation frequencies from ligation reactions involving pDK101 *Xcm*I with *ssr*RNA amplicons were considerably lower, around  $3 \times 10^2$  to  $3 \times 10^3$ . The religation frequency of pDK101 *Xcm*I also fell within this frequency range.

Colonies with recombinant plasmids were selected by  $\alpha$ -complementation on media containing ampicillin and the chromogenic substrate X-Gal. Colonies carrying non-recombinant colonies metabolise the substrate and produce blue colonies, whereas recombinant plasmids with insertions at the TA-cloning site produce white colonies since the *lacZ* coding region is interrupted. As expected, when native pDK101 was transformed into DH5 $\alpha$  cells all the resultant colonies had a blue phenotype. White colonies, which potentially carry recombinant plasmids with *ssr*RNA gene inserts, were evident in all ligations involving *ssr*RNA amplicons. The proportion of white colonies varied from 20% to 52.6%, however, 35.7% of the colonies from the religation transformation were also white. This suggests that a large proportion of the white colonies from the ligation reactions with *ssr*RNA amplicons probably did not contain recombinant plasmids with *ssr*RNA gene inserts. Plasmid DNA was isolated from all the white colonies, digested with *Eco*RV and examined by agarose gel electrophoresis. These colonies only produced plasmid DNA of 3 kbp in size; none of the suspected recombinant plasmids contained a *ssr*RNA gene. A large proportion of the transformed colonies were pale blue in colour, and this colouration persisted when bacteria were subcultured onto fresh media. The phenotype of these colonies was therefore unclear, and the blue-white screening appeared to be ineffective. In subsequent experiments, blue-white screening for recombinant colonies was disregarded and plasmids were only isolated from *ssr*RNA ligations with a transformation frequency five times higher than the religations. However, no plasmids examined possessed the *ssr*RNA gene insert.

**Table 3.1** Transformation of pDK101 TA-cloned PCR-amplified ssrRNA genes into competent DH5 $\alpha$  cells.

Plasmid Species	PCR Product	Tf <sup>1</sup>	Number of Transformed Amp <sup>R</sup> Colonies			Proportion of white colonies
			white (lac <sup>-</sup> )	pale blue (lac <sup>+/+</sup> )	blue (lac <sup>+</sup> )	
pUC18	-	5.0 x 10 <sup>6</sup>	-	-	-	-
pDK101	-	4.8 x 10 <sup>5</sup>	0	0	2400	0%
pDK101 <i>XcmI</i>	- (Religation)	8.0 x 10 <sup>2</sup>	5	6	3	35.7%
pDK101 <i>XcmI</i>	<i>A.polyphaga</i> 2.2 kbp ssrRNA	3.0 x 10 <sup>3</sup>	12	37	3	23.1%
pDK101 <i>XcmI</i>	<i>V.platypodia</i> 1.75 kbp ssrRNA	4.6 x 10 <sup>2</sup>	3	1	4	37.5%
pDK101 <i>XcmI</i>	<i>V.anglica</i> 1.95 kbp ssrRNA	2.1 x 10 <sup>3</sup>	20	16	2	52.6%
pDK101 <i>XcmI</i>	<i>P.mainensis</i> 1.95 kbp ssrRNA	9.1 x 10 <sup>2</sup>	4	8	4	25%
pDK101 <i>XcmI</i>	<i>P.bursella</i> 1.95 kbp ssrRNA	2.9 x 10 <sup>2</sup>	1	4	0	20%

**Table 3.2** Transformation of pCRII TA-cloned PCR-amplified ssrRNA genes into competent One Shot and Maximum Efficiency DH5 $\alpha$ I<sup>Q</sup> cells.

Plasmid Species	PCR Product	One Shot Competent Cells		Maximum Efficiency LacI <sup>Q</sup> Cells	
		Tf <sup>1</sup>	Proportion of white colonies	Tf	Proportion of white colonies
pUC18	-	1.5 x 10 <sup>7</sup>	-	>1 x 10 <sup>8</sup>	-
pCRII	- (Religation)	2.0 x 10 <sup>3</sup>	0%	2.1 x 10 <sup>4</sup>	6.3%
pCRII	550 bp gene	2.8 x 10 <sup>5</sup>	97.1%	2.5 x 10 <sup>6</sup>	95.9%
pCRII	50ng <i>V.anglica</i> 1.95 kbp ssrRNA	2.1 x 10 <sup>5</sup>	40%	1.5 x 10 <sup>6</sup>	90.9%
pCRII	250ng <i>V.anglica</i> 1.95 kbp ssrRNA	1.9 x 10 <sup>5</sup>	33%	1.4 x 10 <sup>6</sup>	87.5%

<sup>1</sup> Transformation Frequency (Tf) = Number of Amp<sup>R</sup> transformants  $\mu\text{g}^{-1}$  of plasmid DNA.

Considering the general low transformation frequency using DH5 $\alpha$  cells, the low proportion of white colonies, the similarity between the transformation results from ligation reactions with *ssrRNA* amplicons and those without, and the absence of any gene insertions from potentially recombinant plasmids, an alternative TA-cloning system was tried.

### **TA-cloning using pCRII**

To improve the chance of successfully cloning *ssrRNAs* using TA-cloning, a commercially available TA-cloning system, which utilises a 3.9 kbp plasmid pCRII, was employed (Invitrogen). To increase the transformation frequency One Shot competent cells, which were included in the TA-cloning kit, and Maximum Efficiency DH5 $\alpha$ F'I<sup>q</sup> cells (Gibco BRL) were used. An amplicon of approximately 550 bp from an unrelated bacterial template was also included to ensure that the TA-cloning was working efficiently.

The TA-cloning results for pCRII with the *Vannella anglica* *ssrRNA* amplification product, and the appropriate controls are shown in Table 3.2. An initial examination of the results clearly demonstrated that the transformation frequencies were considerably higher for the One Shot and Maximum Efficiency DH5 $\alpha$ F'I<sup>q</sup> cells than the Library Efficiency DH5 $\alpha$  cells used previously for the pDK101 cloning. The transformation frequencies for the One Shot cells were  $1.5 \times 10^7$  for pUC18 and between  $1.9$ - $2.8 \times 10^5$  for ligations between pCRII and amplicons. The transformation frequencies for the Maximum Efficiency DH5 $\alpha$ F'I<sup>q</sup> cells were approximately ten-fold higher; over  $1 \times 10^8$  for pUC18 and ranging from  $1.4$ - $2.5 \times 10^6$  for ligations between pCRII and amplicons. Unlike the TA-cloning results with pDK101, the comparative transformation frequency for religation was approximately 100-fold less than ligations with amplicons. Moreover, the proportion of white colonies was far more favourable for pCRII than for pDK101. For pCRII, the proportion of white colonies varied from 33-97.1% for the One Shot cells and 87.5-95.9% for the Maximum Efficiency DH5 $\alpha$ F'I<sup>q</sup> cells for ligations with amplicons, compared to 20-52.6% for pDK101. The proportion of white colonies from religations was also much lower for pCRII varying from 0-6.25% compared to 35.7% for pDK101. Together these results suggested that the pCRII TA-cloning was far more likely to yield recombinant plasmids carrying the amplified *ssrRNA* insert than pDK101.

Since the transformation frequencies, and the proportion of white colonies, was higher for Maximum Efficiency DH5 $\alpha$ F'T<sup>q</sup> cells than One Shot cells, plasmid DNA was isolated from the former, and examined for the presence of a DNA insert. Examination of five plasmids isolated from the control ligation with the 550 bp amplicon identified three recombinant plasmids with an insert of the correct size. Examination of plasmids from the ligation with the 1.95 kbp *V. anglica* *ssrRNA* amplicon suggested that the frequency of white colonies containing this much larger insert was far lower. Examination of plasmid DNA from 110 white colonies (designated pCRV01.1-110) identified only one, pCRV01.106, containing an insert of the correct size. Four other clones were identified which possessed small gene insertions in the range of 300-900 bp. Figure 3.4 A shows plasmid DNA digested with *EcoRI* from clones pCRV01.102-106. All these clones produced a DNA fragment of 3.9 kbp which corresponds to a size of pCRII plasmid DNA. There was no indication of any DNA fragments which have been excised from clones pCRV01.102-105, but in Lane 6 with clone pCRV01.106 a clear DNA band was evident at about 1.7 kbp and a very faint band (arrowed), at low molecular weight was also present.

---

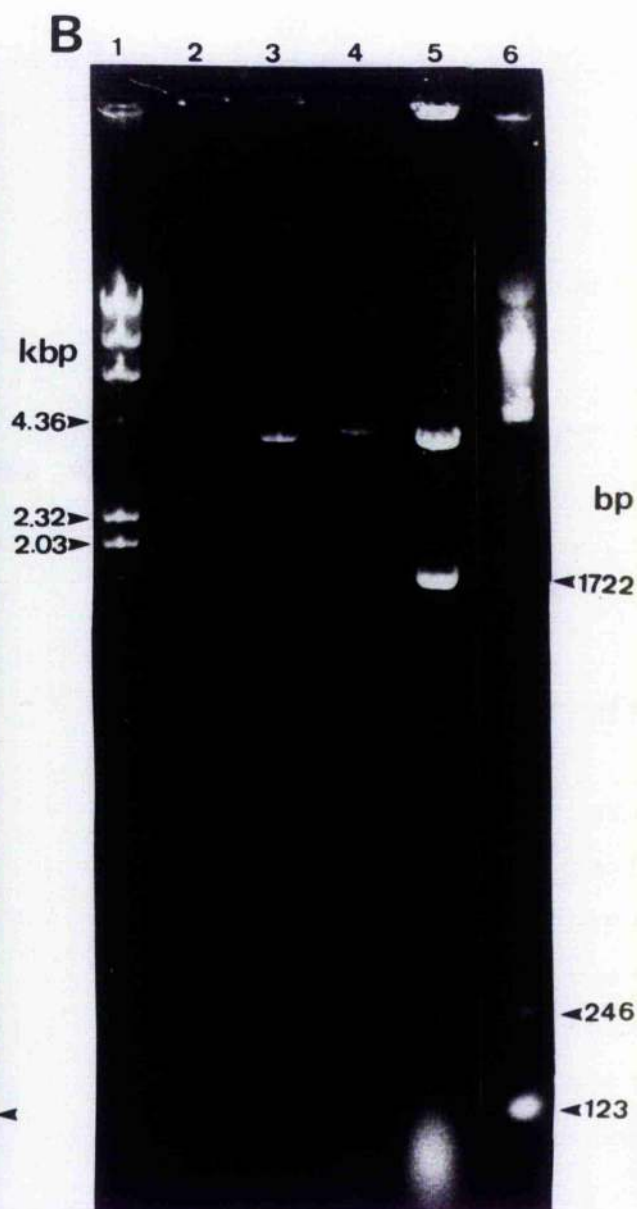
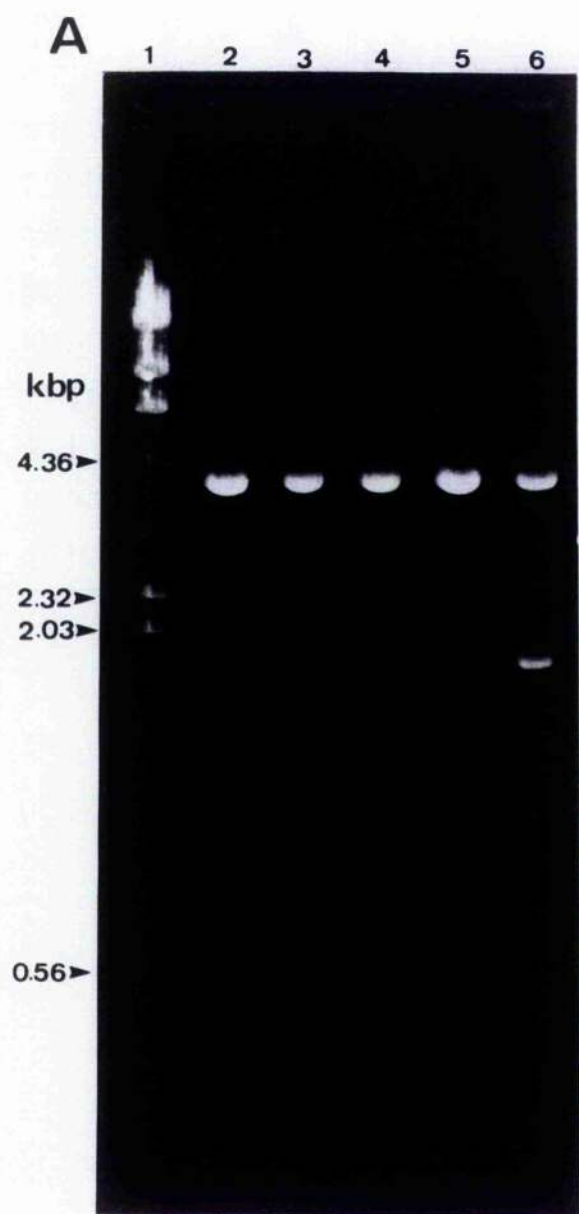
**Figure 3.4** Examination of suspected *V. anglica* *ssrRNA* recombinant clones.

(A).

Lane 1:     $\lambda$  *Hind*III ladder  
 Lane 2:    pCRV01.102 *EcoRI*  
 Lane 3:    pCRV01.103 *EcoRI*  
 Lane 4:    pCRV01.104 *EcoRI*  
 Lane 5:    pCRV01.105 *EcoRI*  
 Lane 6:    pCRV01.106 *EcoRI*

(B).

Lane 1:     $\lambda$  *Hind*III DNA marker  
 Lane 2:    pCRII  
 Lane 3:    pCRII *EcoRI*  
 Lane 4:    pCRV01.106  
 Lane 5:    pCRV01.106 *EcoRI*  
 Lane 6:    123 bp DNA marker



To confirm the existence of the *ssrRNA* insertion, undigested and *EcoRI* digested plasmid DNA from pCRV01.106 and a pCRII religation reaction were compared by electrophoresis (Figure 3.4 B). The undigested pCRII plasmid migrated to the equivalent of 2.5 kbp (Lane 2) compared to approximately 4.2 kbp for the undigested pCRV01.106 plasmid (Lane 4). Because undigested covalently closed circular plasmid DNA migrates more quickly through an agarose matrix than linear DNA fragments of the same size, these plasmids were significantly larger than their mobilities suggested. However, the difference between the two plasmids indicated that pCRV01.106 was in excess of 1.7 kbp larger than pCRII. Digestion with *EcoRI* demonstrated that the size of the pCRII plasmid was 3.9 kbp (Lane 3). For pCRV01.106, in addition to the 3.9 kbp plasmid DNA band, *EcoRI* excised two *ssrRNA* DNA fragments; a large fragment of approximately 1.7 kbp and a small fragment of 230 bp (Lane 5). The combined size of these fragments was very similar to the *V. anglica* *ssrRNA* gene size estimates of 1.95 kbp and 1.98 kbp from the initial amplification of the gene and ribotyping presented in Chapter 2.

Initial sequence data from M13 primers flanking the *V. anglica* DNA insert demonstrated the gene was *ssrRNA* in origin and that the termini of the gene were successfully cloned. The clarity of the *ssrRNA* sequence obtained from pCRV01.106 was reasonably poor, presumably due to the large size of the recombinant plasmid. The *ssrRNA* gene was therefore sub-cloned into more manageable fragments for DNA sequencing.

### **Sub-cloning the *ssrRNA* gene from pCRV01.106**

An initial restriction survey of pCRV01.106 demonstrated that the *ssrRNA* gene did not contain *HindIII*, *KpnI*, *SacI*, *SmaI* or *SpeI* restriction sites, but an *EcoRI* and *BamHI* sites were present. Single and double digestion of pCRV01.106 with *EcoRI* and *BamHI* are shown in Figure 3.5 A. Due to the presence of *EcoRI* sites in the sequence flanking the *ssrRNA* gene, digestion with *EcoRI* completely excised the DNA insert. Lane 2 shows the products of the pCRV01.106 *EcoRI* digestion, with the high molecular weight band of approximately 3.9 kbp corresponding to the vector DNA and two *ssrRNA* gene fragments, one clearly visible at approximately 1.7 kbp, and a faint band of a small fragment of approximately 230 bp (arrowed). In lane 3, the *BamHI* digest produced a high molecular weight band in excess of 4 kbp, and a band of approximately 1.0 kbp. Since only a single *BamHI* restriction site flanks the TA-cloning site in pCRV01.106, this demonstrated that a *BamHI* site



**Figure 3.5** Sub-cloning the *V. anglica* *ssrRNA* gene from pCRV01.106 using *EcoRI* and *BamHI*.

**(A).** Digestion of pCRV01.106 with *EcoRI* and *BamHI*.

- Lane 1:  $\lambda$  *HindIII* DNA marker
- Lane 2: pCRV01.106 *EcoRI*
- Lane 3: pCRV01.106 *BamHI*
- Lane 4: pCRV01.106 *EcoRI BamHI*
- Lane 5: 123 bp DNA marker

**(B).** Sub-cloning the 230 bp *EcoRI* fragment.

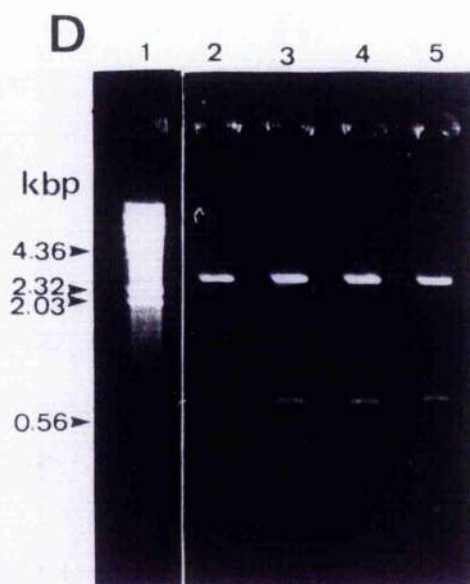
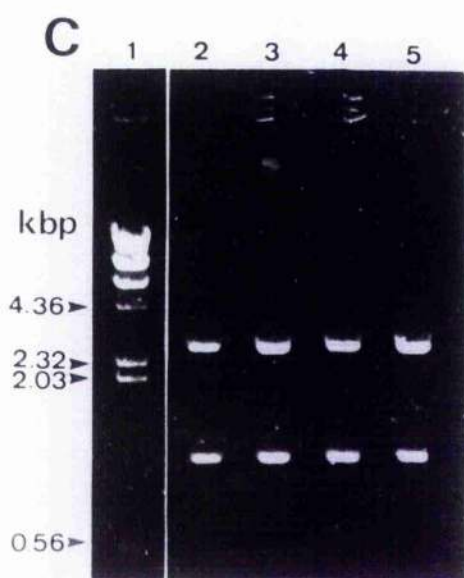
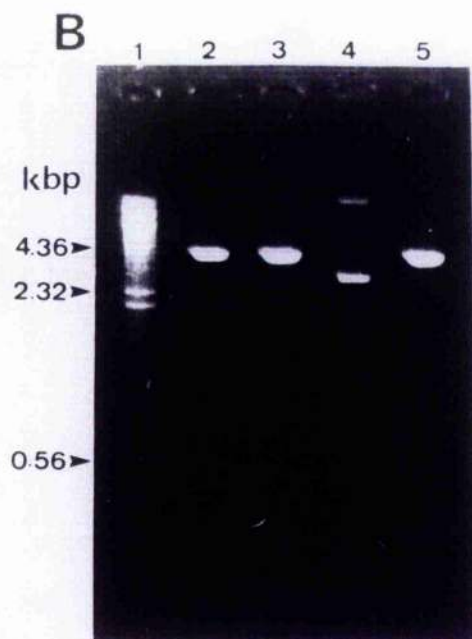
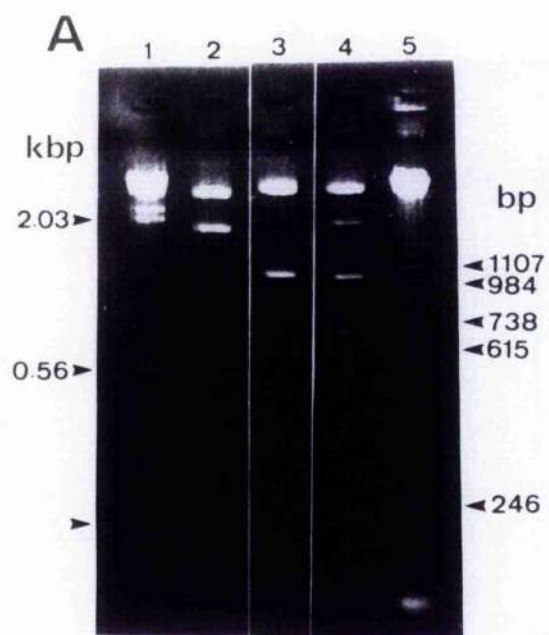
- Lane 1:  $\lambda$  *HindIII* DNA marker
- Lane 2: pCRV03.01 *EcoRI*
- Lane 3: pCRV03.02 *EcoRI*
- Lane 4: pCRV03.03 *EcoRI*
- Lane 5: pCRV03.04 *EcoRI*

**(C).** Sub-cloning the 1 kbp *EcoRI BamHI* fragment.

- Lane 1:  $\lambda$  *HindIII* DNA marker
- Lane 2: pUCV11.01 *EcoRI BamHI*
- Lane 3: pUCV11.02 *EcoRI BamHI*
- Lane 4: pUCV11.03 *EcoRI BamHI*
- Lane 5: pUCV11.04 *EcoRI BamHI*

**(D).** Sub-cloning the 0.7 kbp *EcoRI BamHI* fragment.

- Lane 1:  $\lambda$  *HindIII* DNA marker
- Lane 2: pUCV12.15 *EcoRI BamHI*
- Lane 3: pUCV12.16 *EcoRI BamHI*
- Lane 4: pUCV12.17 *EcoRI BamHI*
- Lane 5: pUCV12.18 *EcoRI BamHI*



is present almost midway along the 1.95 kbp *ssrRNA* gene. The double *EcoRI* *Bam*HI digest, shown in lane 4, produced bands of approximately 3.9 kbp, 1.7 kbp, 1.0 kbp, 0.7 kbp and 230 bp. The largest band corresponds to linear plasmid DNA, the 1.7 kbp fragment is probably a partial *Bam*HI digestion product, and the remaining 1.0 kbp, 0.7 kbp and 230 bp fragments correspond to the complete *EcoRI* and *Bam*HI digestion products. Considering these single and double digestions, and the relative positions of the *EcoRI* and *Bam*HI sites of pCRII, it is clear that the 1.0 and 0.7 kbp fragments are *EcoRI* *Bam*HI fragments, and the 230 bp fragment is an *EcoRI* fragment.

The 230 bp *EcoRI* fragment was ligated with pCRII *EcoRI* and the recombinant colonies were labelled with the prefix pCRV03 followed by the clone number. The 1 kbp and 0.7 kbp *EcoRI* *Bam*HI fragments were cloned into the pUC18, linearised with the same enzymes, and recombinant colonies were labelled with the prefixes pUCV11 and pUCV12 respectively.

Plasmid DNA was extracted from four white transformed colonies for each of the sub-cloning reactions, digested with the appropriate restriction enzyme(s) to excise the *ssrRNA* gene fragment and examined by agarose gel electrophoresis (Figures 3.5 B, C and D). Three of the four pCRV03 clones produced a plasmid DNA band at 3.9 kbp and the expected 230 bp fragment (Figure 3.5 B). Clone pCRV03.3 in lane 4, did not yield a *ssrRNA* gene fragment of the expected size. Since, the restriction pattern suggested that dimerised and covalently closed DNA were present, the plasmid DNA was probably not digested. All four pUCV11, and all four pUCV12 clones produced vector DNA band of 2.7 kbp, and the expected 1.0 kbp and 0.7 kbp *ssrRNA* gene fragments respectively (Figure 3.5 C and D).

## Sequencing

The use of plasmid DNA isolated using Wizard Maxi-prep Systems produced poor quality sequence compared to plasmid DNA isolated using Wizard Mini-prep systems. The absence of RNase in the maxi-prep systems suggests that RNA-free plasmid DNA was required for good sequence resolution. Diligent washing of the precipitated, denatured plasmid DNA, with complete removal of all the supernatant during each step, was also critical for quality sequencing. Denatured DNA stored in solution did not take up the label as well as freshly prepared, denatured and dried DNA which was resuspended directly in the primer solution.

The resolution of the sequence data obtained from plasmids pUCV11 and pUCV12, and to a lesser extent pCRV03, was considerably better than that obtained from the pCRV01.106 template. By using the M13 primers, approximately 50% more sequence could be read from pUCV11 and pUCV12, compared to pCRV01.106. Moreover, DNA obtained from half a plasmid mini-prep was sufficient for a sequencing reaction and with a combination of 2.5 and 4.5 hour gel runs, 350 nucleotides could often be read. The improved DNA sequencing from pUCV11 and pUCV12 was presumably due to the isolation of quality template DNA from these smaller higher copy number plasmids. In most cases, the resolution of sequence data using internal primers was as good as that from the M13 primers, however, sequence data from primer F3 was poor.

### **Primary *ssrRNA* gene sequence**

The amplified *V. anglica* *ssrRNA* gene was cloned and sequenced, and the primary DNA sequence is presented in Figure 3.6. The gene was amplified using the eukaryotic *ssrRNA* primers complementary to conserved sequence elements at the 3' and 5' ends, and found to be 1962 bases in length with a G+C content of 37.1%. Comparisons with other *ssrRNA* sequences suggest that one base may be present distal to the forward primer and up to five bases distal to the reverse primer. This would give the gene a total length of up to 1968 bases. Sequencing could not resolve the nucleotide at position 1071.

Examination of the gene sequence identified all the restriction sites which were found in the sub-cloning in this chapter and the riboprinting in Chapter 2. The *EcoRI* site (G'AATTC) and the *BamHI* site (G'GATCC), which were responsible for producing fragments of approximately 1 kbp, 0.7 kbp and 230 bp, are found at positions 1741 and 1039 respectively. Four *AluI* sites (AG'CT) are present at positions 171, 599, 637 and 1388 producing fragments, in descending order, of 751, 573, 428, 172 and 38 bp. Similarly sized fragments of 800, 585, 415 and 180 bp were estimated from the *V. anglica* *AluI* riboprint. The size of the largest fragment was clearly over-estimated at the expense of the smallest fragment which could not have been detected. Finally, three *HinfI* sites (G'ANTC) at positions 260, 389, 1332 would have produced fragments of 943, 630, 260 and 129 bp. These expected sizes are very similar to the observed *HinfI* fragments of 950, 630, 265 and 140 bp estimated from the *HinfI* riboprint in Chapter 2.

1      ACCTGGTTGA TCCTGCCAGT GATGATATGC TTGAATTAAA GATTAAGCCA  
51      TGCAAAATGTA GTATAAATCA CTTTTACTGT GAAACTGTGA ACGGCATCA  
101      AAATCAGTTA TAGTTTATTTT GATGGTATTG AGGTTTACCT CGAACTACAT  
151      GGATAACTGT AGTAACCTAG **AGCTAATACA** TGCAAAAACA ATTCAACTTT  
201      TTGGAAGAAAT TGGCACTTAT TAGATATCAA ACCAATATAT TTTCGGATAT  
251      TTTTATGGTG **AGTC**ATAATA ACTGAACAGA TCGCATGACC TCGTGTCCGC  
301      GACGTTTCAT TCAAAATTTCT GCCCTATCAA CTAGATGTTA GTATATTGGA  
351      CTAACATGGT GGTAACGGGT AACGGGGAAT TGGGGTT**CGA** **TTCCGGAGAA**  
401      GGAGCATGAG AAACGGCTAC TACTTCTAAG GAAAGCAGCA GCGCGCAAA  
451      TTATTTCAATG GAAACAATTT TCCGAAATAG TGACAATTAA TAACAATACG  
501      TTGCATAAAT TGTGATGTAA TTGGAATGAG TACAGTTTAA AACCCTTAAC  
551      GAGTAACAAAT TAGAGGGCAA GTCTGGTGCC AGCAGCCGCG GTAAAT**CCAG**  
601      **CTCTA**ATAGT GTATATCTAT GTTGTTCAG TTAAAA**AGCT** CGTAGTTGAA  
651      TTTTGAATAT TTAAATAATG TTTGGTTTACT TTTTATAAGTA TGA CTACTTA  
701      TTTTATATCCT TCTTACTAGT AACTCGATCT GTTCTTAATT GAATGGAATA  
751      GCATCTTTTCG GCGGTGTCCA TAGGATGTAT ATCTCCTCCA ATAAAGTGGAG  
801      AGCGGTTAAA ACCTTAAAGT TTATTGACTG CATCAACATA TATATTTTAA  
851      TTGGAGTTCT AGTTCAATTA CTTTGAGAAA ATTAGAGTGT TTAAGGCAAC  
901      CTCTTGGCCT TCTCAATACA TTAGCATGGG ATAATAAAAT AGGATTTAAA  
951      TTATATATTT GTTGGTTTGA ATAAATTTAAA TAAATGATTA TGGGATAGT  
1001      TGGGGGGATTT TATATTTGGA GTTTAGCGGT GAAATGCATG **GATCCATCA**  
1051      AGATAAACTA AAGCGAAAGC NTTTACCAAG GATGTTTCA TTAAATCAAGA  
1101      ACGAAAGTTA AGGGATCGAA GATGATCAGA TACCGTCCGA GTCTTAACTA  
1151      TAAACTATGC CGACCAGGA TATAAGTTCT TATGAATATC CTTTTGTGTA  
1201      TTCAATTTCC TCTTGTGGAC CTTGTGAGAA ATCAATAAGTT TTTGGGTTCT  
1251      GGGGGGAGTA TGGTCCGAAG GCTGAAACTT AAAGGAATTG ACGGAGGGGC  
1301      ACAACCAGGA GTGGAGCGTG CGGCTTAATTT **TGACTCAACA** CCGGAAAACCT  
1351      TACCAGGTCC AGACATTAGA AGGATTGACA GATTGAT**AGC** **TCCTTCATGA**  
1401      TTTAATGGGT GGTGGTGCAT GCGCGTTCTT AGTTGGTGGT GTGATTTGTC  
1451      TGGTTAATTC CGATAACGAA CGAGACCTTT ACCATACATA ATAGTTTGAT  
1501      CTATTCCTTAT TGTATAGATT TATTTTAAA CTCTTAGAG GGACTATTGG  
1551      TTTCTAAACC AAAGGAAGTA AAGGCGAATA CAGGTCGTG ATGCCCTTAG  
1601      ATGTTCTGGG CTGCACGCGC GCTACAATGA TAATCTCAAC GAGTATTTAT  
1651      TTTATTATAT TGCTTTATTG GAAAATATGG CTAATCTTAA ACGATTATCG  
1701      TGCTTGGGAT AGATGATTGC AATTATTTAT CTGACGAG **GAATTC**CTAG  
1751      TAATCGTATA TCATCAGTGT GCGATGATTA CGTCCCTGCC CCTTGTACAC  
1801      ACCGCCCCGC GCTGCTACCG ATTGGATGGT CCGGTGAAAT CCTCGGATAA  
1851      CAATAAATAC TTTACTTTTGT AAAGTTTATT GTTAGAAGTT GATTAACCT  
1901      TATCATCTAG AGGAAGAAAA AGTCGTAACA AGGTTCCGT AGTCAACCT  
1951      GCAGATGGAT CA

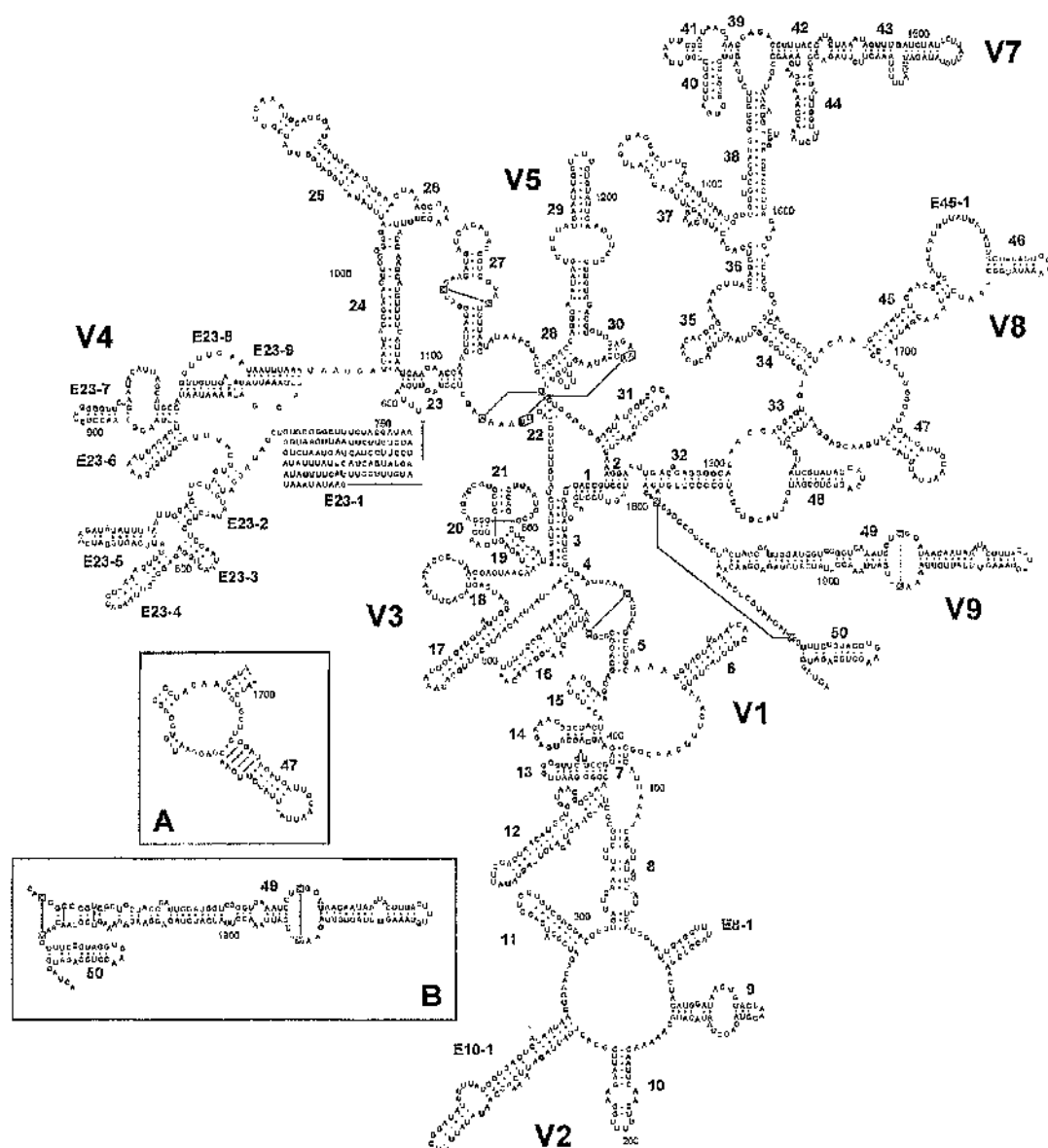
**Figure 3.6** Primary DNA sequence of the *V. anglica* *ssrRNA* gene. The non-coding DNA strand is given in a 5' to 3' direction. The forward and reverse primers used for amplification of the gene are underlined, and the *EcoRI*, *BamHI*, *AluI* and *HinfI* restriction sites are shown in bold.

## Secondary structure of the *ssrRNA*

From comparisons of the predicted secondary structures of other eukaryotic *ssrRNAs*, the secondary structure of the *V. anglica ssrRNA* was determined (Figure 3.7). The predicted secondary structure was typical of a eukaryotic *ssrRNA*. There were no significant differences in the secondary structure of the conserved regions compared to other eukaryotic *ssrRNAs*, and variations which occurred in the variable regions did not conflict with the generalised secondary structure model. All 50 universal *ssrRNA* helices were identified, and the eukaryotic specific E10-1, E23-2, E23-5, E23-6, E23-7 and E23-9, which occur in all eukaryotes except some ancient eukaryotic taxa lineages, were also present. Although the location of helix E23-1 was obvious, the secondary structure of this eukaryotic specific helix, which accommodates approximately 110 nucleotides, could not be unambiguously determined. The RNA secondary structure prediction program FOLDRNA showed that E23-1 may be composed of two adjoining helices, but two large internal loops and an asymmetrical bulge and some mismatched pairings do not lend much support. The structure of the 'helix' E23-1 therefore remains uncertain and the nucleotides for this region are simply listed in Figure 3.7. In addition to the universal eukaryotic helices, helices E8-1, E23-3 and E23-4, which are only exhibited in a minority of eukaryotic taxa, were also predicted for the *V. anglica ssrRNA*. Helices E10-2 and E23-10 were not present.

The presence of helix E8-1 is a relatively rare feature among *ssrRNAs*. However, a composition of a total of six consecutive canonical and G:U base-pairings, and a uracil rich hairpin loop was strong evidence for the existence of this helix. A short helix composed of four consecutive canonical pairs was predicted for helix E23-3, and a longer helix, composed of eight complementary pairs and perhaps one A:G non-canonical pairing, was predicted for helix E23-4. Both these helices extend from a bulge which also connects helices E23-2 and E23-5.

Other interesting features determined for the *V. anglica ssrRNA* secondary structure included a bilateral bulge in helix 29 of the V5 region, a unique internal loop in helix 43 of the V7 region, and an expansion segment between helices 45 and 46 of the V8 region. In most *ssrRNAs*, helix 29 forms a long stem, but in this case a significant bilateral bulge in the middle of the helix was evident. The high frequency of cytosine and uracil residues within the bulge provides strong evidence for its authenticity; it was not a consequence of one or two misincorporated bases.



**Figure 3.7** Predicted secondary structure of *V. anglica* ssrRNA. The helices are numbered in order of occurrence from the 5' terminus. Eukaryotic specific helices are prefixed with an 'E'. Remote interactions are represented by boxed nucleotides joined by unbroken lines. Inserts A and B represent the alternative structures for helices of 47 and 49, respectively. These helices have been extended by interactions involving non-canonical pairs which are indicated by connecting lines.

The internal loop in the middle of helix 43 was also of interest. The eight nucleotide loop was rich in uracil residues which ensures that no interactions occur between bases of this loop. A similar uracil rich region was also found in the small expansion segment between helices 45 and 46. In some *ssrRNAs*, the expansion segment at this site forms a helix which is denoted E45-1, however, in this instance the high proportion of uracil residues ensures that a helix could not be formed.

Beyond the prediction of recognised helical elements, higher-order interactions which encompasses pseudoknots, lone-pair interactions and the involvement of non-canonical base-pairing within helices and as possible extensions of helices were also considered.

### **Pseudoknots**

Previous models of eukaryotic *ssrRNA* secondary and higher-order structure have identified four pseudoknots. All four pseudoknots were found in the *V. anglica* *ssrRNA*, and are shown in Figure 3.7. The first pseudoknot which formed helix 2, involved four canonical pairs from the hairpin loop of helix 1 and the region between helices 31 and 32. The second pseudoknot which formed helices 20 and 21 was G+C rich; seven of the eight interactions involved G:C base-pairings. The composition of the third pseudoknot, which formed helices E23-8 and E23-9 was quite different. Of the 19 interactions, there were 17 A:U interactions, two non-canonical G:U interactions and a single unilateral bulge involving an adenine residue. The final pseudoknot involved interactions which were separated by other secondary structure elements and could not be easily represented. The two nucleotide pairs involved in this interaction were between the hairpin loop of helix 30 and the bulge connecting helices 22, 23, 27 and 28. In Figure 3.7, the nucleotides involved with this interaction are boxed and connected by a solid line.

### **Lone-pair interactions**

Five interactions between lone-pairs were detected which were consistent with previously proposed higher-order interactions. There was a canonical G:C pairing within the bulge which connects helices 4, 5 and 16. Similarly, a second lone-pair interaction involved an A:U base-pairing in the bulge which connects helices 22, 23, 27 and 28. Lone-pair interactions were also detected in the bulges of helix 27 and helix 49, both of which involved G:C pairs. The final interaction also involved a G:C pair located within the stretch immediately before and after helix 49.



### **Non-canonical base-pairing**

Non-canonical G:U base-pairing occurs commonly in RNA secondary structure, but other classes of non-canonical base-pairing are far less frequent. Throughout the predicted secondary structure there were a total of 33 potential non G:U, non-canonical base-pairing within helices. The G:A class were the most frequent, with potentially 14 pairs, four of which were in helix 24, two in helix 49 and single pairs in helices 4, 9, 14, E23-2, E23-4, E23-5, 25, and 29. There were seven potential U:U pairs located within helices 3, 8, 36, 43 and 46, 50, six A:A pairs within helices 4, 10, 23, 44 and 49, three C:A pairs within helices 24, 43 and 50, two C:U pairings in helices 49 and one G:G in helix 23. No potential interactions were detected between pairs of cytosines.

### **Non-canonical extended helices**

Helices 47 and 49 may be extended by the formation of non-canonical base-pairs. The alternative structures for these helices are shown in Figure 3.7 (Inserts A & B). Covariation in the sequence immediately before and after helix 47 suggests that the helix may be extended by four non-canonical pairings; a U:U followed by three G:A pairings, and terminated by at least one, perhaps as many as four canonical or G:U interactions. The extended version of helix 49 is a little more complex. This helix may be extended by six interactions; four G:C base-pairs and two non-canonical interactions which includes a U:U and a A:C. Unlike the extension of helix 47, these interactions did not run consecutively, they were interrupted by two small asymmetrical bulges.

## Discussion

### **TA-cloning with pDK101**

Cloning with pDK101 was unsuccessful. The low frequency of transformation was a general problem, but the similarity of the transformation frequencies between religations and ligations with *ssrRNA* gene amplifications, and the general low proportion of white colonies were perhaps more telling indicators that cloning *ssrRNA* genes with pDK101 was ineffective. The equivalent transformation frequencies suggested that all the transformed colonies resulting from the ligation between linear pDK101 and *ssrRNA* amplicons were pDK101 religations, and examination of plasmid DNA demonstrated that this was the case.

In most instances the proportion of white colonies was low with all ligations except one being less than 50%. The majority of the transformed colonies were blue or pale blue, and these may have arisen in a variety of ways. Undigested pDK101 may have been carried over during the isolation of linearised pDK101, but it is more likely that semi-digested pDK101 was carried over. Since the two *Xcm*I sites are only 15 bp apart, pDK101 digested at only one of the *Xcm*I sites would not be effectively separated by electrophoresis from pDK101 digested at both sites. Due to the close proximity of the *Xcm*I restriction sites, and the consideration that the *Xcm*I recognition site covers 15 bp, cleavage at these sites is probably not independent, and competition may favour cleavage at one particular site. Moreover, cleavage at the termini of DNA fragments may be difficult with some restriction endonucleases (Kaufman & Evans, 1990). Therefore, the carry-over of pDK101 digested at a single *Xcm*I, and religation is a distinct possibility. A low proportion of white colonies, and the presence of both blue and pale blue colonies has occurred in previous studies with TA-vectors cut with *Xcm*I (Kovalic *et al.*, 1991; Mead *et al.*, 1991). Sequence analysis of the cloning junction has shown that pale blue colonies may arise when the translation reading frame is out of register due to small deletions (Lobet *et al.*, 1989; Mead *et al.*, 1991). Alternatively, blue colonies may have arisen from the religation of incompatible deoxythymidylate overhangs (Wiaderkiewicz & Ruiz-Carrillo, 1987). White colonies with non-recombinant plasmids probably arose from the removal of the 3' deoxythymidylate overhangs and blunt-end ligation.

Previous studies using *Xcm*I TA-cloning vectors have been used to clone efficiently small amplicons less than 1 kbp in length (Kovalic *et al.*, 1991; Mead *et al.*, 1991; Harrison *et al.*, 1994). However, there is only a single report of a fragment larger than *V. anglica* *ssrRNA* gene being cloned and this report acknowledges the difficulty of cloning larger products (Mead *et al.*, 1991). Moreover, low transformation frequencies and poor blue-white selection can occur when cloning even small fragments into pDK101 (Westrop, pers. comm).

### **TA-cloning with pCRII**

The One Shot and Maximum Efficiency DH5 $\alpha$ I<sup>q</sup> cells, used for pCRII cloning, were far more competent than the Library Efficiency DH5 $\alpha$  cells used for pDK101. The transformation frequency for pCRII and either the *V. anglica* *ssrRNA* or the 550 bp control amplicons were approximately 100-fold higher than the pCRII religation, and yielded a much higher proportion of white colonies. Together these results suggested that the majority of the white colonies contained recombinant plasmids. However, examination of 110 white colonies from the *V. anglica* *ssrRNA* ligation only identified one clone which possessed the *ssrRNA* gene insert. Since the transformation frequency and the proportion of white colonies were much lower for the religation reaction, these results suggest that the low frequency of recombinant plasmids was not due to the removal of the single 3' deoxythymidylate overhangs or mismatch base-pairings between incompatible plasmid deoxythymidylate overhangs.

Examination of five white colonies from the control ligation with the 550 bp insert, identified three which contained a fragment of the correct size which demonstrating that TA-cloning with pCRII was working effectively. It appeared that cloning the 1.95 kbp *ssrRNA* gene amplicon was considerably more difficult. Plasmid pCRII has been used to clone the amplified *ssrRNA* gene from the acetospora *Haplosporidium nelsoni* (Fong *et al.*, 1993), although no details were given regarding the cloning efficiency.

Plasmids carrying the *ssrRNA* may have been toxic to the cell or unstable resulting in the selective loss of the *ssrRNA*. To examine this possibility, Maximum Efficiency DH5 $\alpha$ I<sup>q</sup> cells were used which over-express the Lac repressor protein preventing any background expression of the *ssrRNA* gene from the Lac promoter. However, there was no difference in the transformation frequency between transformed cells

grown on ampicillin media with or without the Lac inducer IPTG, which suggests that the background expression of the *ssrRNA* gene was not toxic. Furthermore, if recombination was taking place resulting in the selective loss of the *ssrRNA* from the plasmid, a series of clones may have been expected with small gene fragments, but the competent cells had *recA1* genotypes which reduce recombination, and only 4/110 clones were detected with smaller insertions. Therefore, inhibition of the cell or selective loss of the complete *ssrRNA* gene appeared to be unlikely.

Using an identical technique, partial *ssrRNA* gene amplicons of 400 bp in size were cloned with ease (Chapter 5). Despite lower transformation frequencies and lower proportions of white colonies, 100% of the plasmids examined from white colonies contained an insert of the correct size. Therefore, the difficulty encountered cloning the complete *ssrRNA* gene is probably a function of the size of the gene. During the TA-cloning of the complete *ssrRNA* gene, the white colonies may have resulted from ligation of the plasmid with small undetectable DNA fragments, removed from the purified *ssrRNA* amplicon; fragments of this size would out compete the large 1.95 kb *ssrRNA* amplicon at the TA-cloning site.

### **Composition of *V. anglica* *ssrRNA***

The *V. anglica* *ssrRNA* sequence gene sequence obtained from the cloned gene fragments was 1962 nucleotides in length. Comparisons of the sequence with the size of the amplified *ssrRNA* product and the riboprinting results indicated that the complete gene was successfully sequenced, excepting as many as six nucleotides which were distal to the amplification primers. The predicted secondary structure confirmed that essentially the complete gene has been sequenced and that intron sequences were not present. Introns have been identified in some eukaryotic groups including algae, fungi, amoeba-flagellates and the naked amoeba genus *Acanthamoeba* (Wilcox *et al.*, 1993; De Jonckheere, 1993; Gast *et al.*, 1994a; Egger *et al.*, 1995), but no intron was detected in the *V. anglica* *ssrRNA*. Similar *ssrRNA* sizes of other *Vannella* species and members of the morphologically similar genus *Platyamoeba* from Chapter 2, suggests that the absence of introns may be a common feature among this type of naked amoebae.

Examination of the composition of the *ssrRNA* gene sequence demonstrated that the G+C content was comparatively low at 37.1%. A majority of eukaryotic *ssrRNAs* have a G+C content of between 42-50% (Maidak *et al.*, 1996). However, the range

of G+C compositions varies considerably among the eukaryotic ssrRNAs from as low as 28% in the insect *Xenos vesparum* (Chalwatzis *et al.*, 1994) to as high as 75% in the diplomonad *Giardia lamblia* (Sogin *et al.*, 1989). The low G+C content of the *V. anglica* ssrRNA can be attributed to the composition of the variable regions. The average G+C content for the conserved regions was 42.7% compared to just 29.7% for the combined variable regions. Examination of each variable region showed that the G+C composition ranged from 22.9% in the V8 region to 35.6% in the V2 region, which indicates that all the variable regions contributed to the overall low G+C content, a characteristic common to ssrRNAs with low G+C contents (McCutchan *et al.*, 1988; Que & Reed, 1991).

### **Secondary structure and higher-order interactions**

The predicted secondary structure was characteristic of other eukaryotic ssrRNAs. All the universal eukaryotic helices have been determined and helices E8-1, E23-3 and E23-4, which occur exceptionally in some eukaryotic taxa, were also revealed. All four pseudoknots and all five lone-pair interactions, which have been previously proposed (Gutell *et al.*, 1985, 1986; Woese & Gutell, 1989; Neefs & De Wachter, 1990; Gutell, 1993), were also present. There were no regions present that did not conform to the generalised eukaryotic ssrRNA model (Neefs *et al.*, 1993), however there were several features of secondary structure and higher-order interactions worthy of discussion.

#### **Helix E8-1**

The presence of a helix E8-1 is a relatively rare occurrence, but the six consecutive base-pairs clearly indicated its presence in the *V. anglica* ssrRNA. This helix is not present in any other naked or testate amoebae (Gunderson & Sogin, 1986; Weekers *et al.*, 1994; Bhattacharya *et al.*, 1995). This helix has been determined in the ciliate *Euplotes aediculatus*, the euglenoid *Euglena gracilis*, some kinetoplasts (De Rijk *et al.*, 1992), and most spectacularly in the plasmodial slime mold *Physarum polycephalum* which has a large G+C rich helix of 13 consecutive base-pairs (Johansen *et al.*, 1988).

#### **Helices E23-1 to E23-5**

A satisfactory structure could not be determined for 'helix' E23-1, although the structures for helices E23-2, E23-3, E23-4 and E23-5 were established. The secondary structure prediction of the V4 region has been notoriously difficult to

determine (Neefs & De Wachter, 1990; Nickerent & Sargent, 1991). Some previous reports have completely avoided examining the secondary structure of the V4 region (Gutell *et al.*, 1985; Hinkle *et al.*, 1994), whereas others have attempted to decipher only part of the V4 secondary structure (Gutell, 1993). Since the discovery of the pseudoknot in the V4 region, most problems have been associated with the region between helices 23 and E23-6 (Neefs & De Wachter, 1990). In general, the secondary structure of this region is difficult to determine because sequence conservation is very low. Consequently, as the number of sequences available for comparative purposes has increased, the secondary structure model for this region has altered (Neefs & De Wachter, 1990; Neefs *et al.*, 1991; De Rijk *et al.*, 1992). However, difficulties resolving the secondary structure of this region persist, particularly if it is large or has a base composition bias (Kwon *et al.*, 1991). Unfortunately, this is the case for the *V. anglica* ssrRNA; the whole of the V4 region is A+U rich, and helix E23-1 alone covers in excess of 100 nucleotides.

The secondary structure of helices E23-2, E23-3, E23-4 and E23-5 is consistent with the current secondary structure model (De Rijk *et al.*, 1992; Neefs *et al.*, 1993). Most eukaryotic ssrRNAs possess helices E23-2 and E23-5, but helices E23-3 and E23-4 are less common. Only platyhelminths, insects, and the euglenoid *E. gracilis* have both these helices. Helix E23-4 is present in the amoeba-flagellate *N. gruberi* and the naked amoeba *Acanthamoeba castellanii* (Neefs *et al.*, 1993). However, helices E23-3 and E23-4 are not present in any other naked or testate amoebae ssrRNAs (Weekers *et al.*, 1994; Bhattacharya *et al.*, 1995). The conformation of helices E23-2, E23-3, E23-4 and E23-5 of *V. anglica* is similar in *Drosophila melanogaster*, despite the apparent lack of sequence conservation. The lengths of helices are similar, although helix E23-3 of *V. anglica* is approximately half the size of *D. melanogaster* (De Rijk *et al.*, 1992).

### **Bilateral bulge of helix 29**

Helix 29, which makes up the V5 variable region, is also rather unusual in *V. anglica*, due to the presence of a comparatively large bulge in the middle of the helix. Single unilateral bulges and rare classes of non-canonical base-pairings are common in this helix, but larger bulges are usually apparent at the base of the helix in the case of chordates (Gutell, 1994). However, helix 29 of the *V. anglica* ssrRNA is considerably larger than most other ssrRNAs, and the high proportion of uracil and cytosine residues indicates that the internal bulge is present, as several

mismatches would be required for complementary base-pairing.

### **Internal loop of helix 43**

The uracil rich internal loop, which is located five base-pairs distal to the two consecutive mismatched pairs in helix 43, is a structural feature unique to the *V. anglica* ssrRNA. However, given the structural variation at this site in helix 43 this is not surprising. Comparisons with other ssrRNAs demonstrate that four to six base-pairs distal to the conserved mismatched bases, there may be a single unpaired nucleotide, a symmetrical bulge, or a small loop. A single unpaired base is found in the apicomplexa *Toxoplasma gondii*, the chlorophyte *Chlorella luteoviridis*, the molluscs *Placopecten magellanicus* and *Onchidella celtica*, and the echinoderm *Strongylocentrotus purpuratus* (Gutell, 1994; Winnepenninckx *et al.*, 1994; Gagnon *et al.*, 1996). Symmetrical bulges are present in the apicomplexa *Babesia bigemina*, and the chordates *Homo sapiens* and *Mus musculus* (Gutell, 1994), whereas small internal loops are also found varying in size from two nucleotides in the ciliate *Oxytrichia nova* to five nucleotides in the rhodophyte *Palmaria palmata* (Neefs *et al.*, 1993; Van de Peer *et al.*, 1994). Comparisons of partial ssrRNA sequences from other morphologically similar amoebae (Chapter 5) suggests that an internal loop, located at an identical position, is present in other *Vannella* and *Platyamoeba* species. From the ten other isolates examined, seven had internal loops similar to *V. anglica* with an A+U rich loop of six to ten bases in length. The loop of *P. plurinucleolus* was a little more unusual as it was 16 nucleotides in length. The internal loops *V. lata* and *V. simplex* were only three nucleotides in length, which is more characteristic of other ssrRNAs.

### **'Helix' E45-1**

The region joining helices 43 and 44 in *V. anglica* was 17 nucleotides in length and composed entirely of adenine and uracil bases. Comparisons of other eukaryotic ssrRNAs indicate that this region is usually four to seven nucleotides in length with a reasonably even base composition (Gutell *et al.*, 1985; De Rijk *et al.*, 1992; Neefs *et al.*, 1993; Gutell, 1994; Van de Peer *et al.*, 1994, 1996). However, there are a few notable exceptions. For the naked amoebae *Phreatamoeba balamuthi*, *A. castellanii*, *Entamoeba histolytica* and the apicomplexa *Plasmodium falciparum*, this region is approximately 80, 65, 65 and 55 nucleotides respectively (Gunderson & Sogin, 1986; McCutchan *et al.*, 1988; Que & Reed, 1991; Hinkle *et al.*, 1994). The base composition of *P. balamuthi* and *A. castellanii* are reasonably well balanced and

they form a single helix (Sogin & Gunderson, 1987; Hinkle *et al.*, 1994), which is denoted E43-1. For *E. histolytica* and *P. falciparum*, the composition of this region is heavily biased in favour of adenine and uracil residues. The secondary structure of this helix has not been determined for *E. histolytica*, and although *P. falciparum* does form a helix, there is a substantial bulge at the base and two smaller internal bulges as a result of the high frequency of uracil residues (Gutell, 1994).

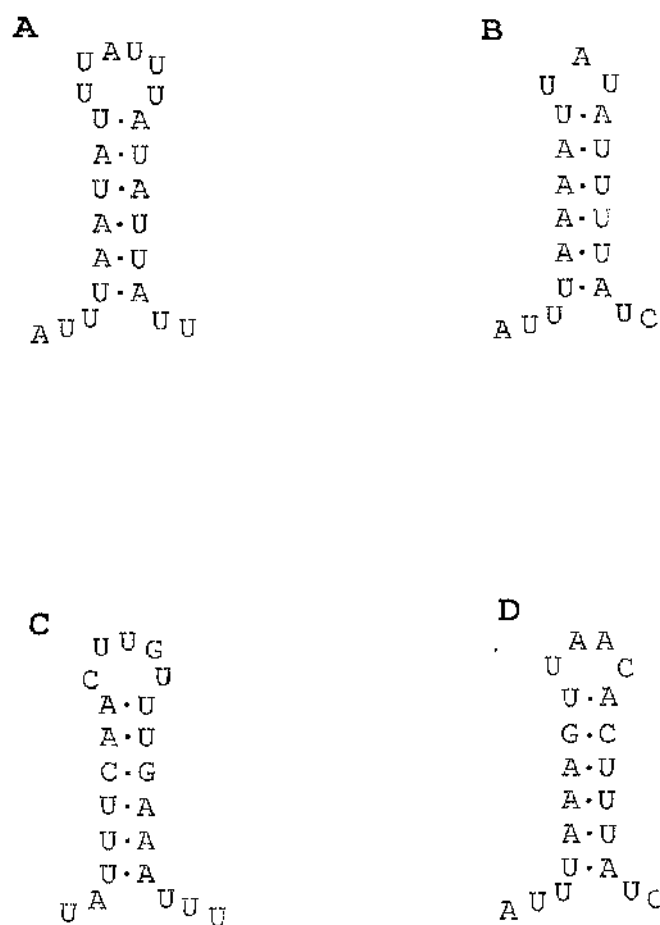
Examination of the partial *ssrRNA* sequences, presented in Chapter 5, demonstrates that other *Vannella* and *Platyamoeba* isolates are similar to *V. anglica* in this region; all are A+U rich and size ranges from 13 to 23 nucleotides. Examination of the secondary structure indicates that some isolates may form a small helix consisting of six consecutive canonical base-pairs. The predicted E-45.1 helices for *V. arabica*, *V. septentrionalis*, *V. lata* and the *Vannella*-like VP3 isolate are shown in Figure 3.8.

The interesting secondary structure features so far described are all located within variable regions of the *ssrRNA*. It was thought that variable regions were unlikely to have any major structural or functional importance. All the critical functional sites, inferred from experimentation on prokaryotic *ssrRNAs*, are located within conserved regions of the molecule (Raué *et al.*, 1990). However, eukaryotic systems are being developed for the direct examination of the functional sites of *rRNAs* in *Tetrahymena thermophila* and yeast (Sweeney *et al.*, 1993; Venema *et al.*, 1995). A recent study using the *Tetrahymena* system indicated that a variable region in the large-subunit ribosomal RNA was essential for either RNA processing or stabilization of the mature form. Replacement experiments indicated that the secondary or tertiary structure of the region provided the essential function, not the primary sequence or the size (Sweeney *et al.*, 1994). Further studies of this sort should provide valuable insight into the functional role of the structural elements in the variable regions of the ribosomal RNAs.

### **Pseudoknots**

The *V. anglica* *ssrRNA* possessed the four pseudoknots and the three lone-pair interactions which have been previously predicted (Gutell *et al.*, 1986; Woese & Gutell, 1989; Neefs & De Wachter, 1990; Gutell, 1993). The sequence composition of the two pseudoknots which form helix 2, and helices 20 and 21 is highly conserved, and they are thought to have important structural and functional roles.





**Figure 3.8** Predicted secondary structure of *ssrRNA* E43-1 helices. (A) *V. arabica*, (B) *V. septrentionalis*, (C) VP3 and (D) *V. lata*.

Site-directed mutagenesis of these regions in *E. coli* indicates that the former is essential for the stability of the ribosome (Poot *et al.*, 1996), and the latter is essential for ribosome activity (Powers & Noller, 1991). Moreover, the pseudoknot involved in the long distance interaction between the bulge connecting helices 22, 23, 27 and 28 and the hairpin loop of helix 30, is indispensable for translation (Vila *et al.*, 1994).

The remaining pseudoknot forms the helices E23-8 and E23-9 in the V4 region. Although this pseudoknot is present in essentially all eukaryotic ssrRNAs, sequence conservation is poor. Helices E23-8 and E23-9 in the *V. anglica* ssrRNA are A+U rich with 17 of the 19 base-pairs between adenine and uracil bases. Comparisons of this region in other ssrRNAs indicates that helix E23-8 is often A+U rich, but helix E23-9 usually has a more even base composition (De Rijk *et al.*, 1992; Neefs *et al.*, 1993; Van de Peer *et al.*, 1994, 1996). Since the other three pseudoknots have structural or functional importance, it is tempting to speculate that this extensive pseudoknot is also important. The variable sequence composition would suggest that higher-order structure of this pseudoknot is more significant than a specific sequence element.

### **Non-canonical extended helices**

From comparative sequence analysis it has been suggested that helices 47 and 49 may be extended by non-canonical base-pairing (Gutell *et al.*, 1994). The alternative form of helix 47 in *V. anglica* is extended by four unusual non-canonical pairs, a G:U pair and perhaps two canonical pairs. The non-canonical pairs include a U:U pair and three G:A pairs. This combination of non-canonical pairs is proposed for the extended helix 47 of a majority of eukaryotic ssrRNAs, however some others have a C:A pair instead of the first G:A pair, and the green alga *Chlorella luteoviridis* has a C:C pair at this site (Gutell *et al.*, 1994). Comparisons of sequence data for this region from morphologically similar isolates, presented in Chapter 5, also support the extension of helix 47 by non-canonical pairings. Sequence data from 11 other isolates indicates that nine isolates possess this common motif of one U:U pair, three G:A pairs and one G:U pair. Species E has a C:A pair in place of the first G:A pair, and *V. simplex* is a little unusual as it starts with a C:U pair, instead of the usual U:U pair and is followed by an A:A pair instead of a G:A pair. There is no support for canonical base-pairings extending beyond the G:U pair.

The extended version of helix 49 of *V. anglica* is more complex than the extended version of helix 47. The extended region is composed of six interactions, two of which involve non-canonical base-pairs, but the interactions are not consecutive; they are interrupted by two small asymmetrical bulges. Despite non-canonical pairings and asymmetrical bulges, the extended helix 49 is characteristic of other *ssrRNAs* (Gutell *et al.*, 1994). The apparent weak associations in the extended region of this helix may be functionally important, as bases in this region have been implicated in elongation in *E. coli* (Raué *et al.*, 1990).

### **Non-canonical pairings**

Comparisons of secondary structures and tertiary interactions in other *ssrRNAs* indicate that the majority of the potential non-canonical pairings detected in regular helices of the *V. anglica ssrRNA* are apparent. The non-canonical G:A pairs in helices 9, 14, 24, 25 and 29 have been previously predicted (Gutell *et al.*, 1994). The potential U:U pairs in helices 8, 36 and 43, the A:A pairs in helix 10, 23 and 49, the C:A pairs in helix 43, the G:G pair in helix 23, the C:U pair in helix 49, are all common, although whether or not these potential pairs interact remains unclear. The C:A pair in helix 24, the A:A pair in helix 44 and the U:U pair in helix 46 occur infrequently or are located at particular sites where mismatches and small bulges are frequent, so these potential pairs probably do not interact but are present nevertheless. The potential G:A pairs in helices E23-2, E23-4 and E23-5 cannot be verified, because sequence conservation within these helices is poor, and mismatches, particularly of the G:A class, are not uncommon in this region. The six remaining potential non-canonical pairs, the G:A pairs in helix 4 and 49, the U:U pair in helix 3 and 50, the A:A pair in helix 4 and the C:A pair in helix 50, occur in conserved regions of secondary structure. These may be natural mismatches, but they may also be a result of a sequence error.

### **Fidelity of the *ssrRNA* sequence**

The *V. anglica ssrRNA* was amplified by the polymerase chain reaction, cloned into a TA-cloning vector, sub-cloned and sequenced. During the PCR process, thermostable *Taq* DNA polymerase misincorporates bases in a template independent fashion at a low, but measurable frequency. The rate of misincorporation depends on a variety of reaction conditions such as the annealing temperature, the concentration of nucleotides and magnesium chloride and the number of cycles (Innis & Gelfand, 1990). Under normal reaction conditions a cumulative error

frequency of up to 0.25% can be expected for 30 cycles of amplification (Saiki *et al.*, 1988). Misincorporated nucleotides can be identified by sequencing independently isolated clones. In this chapter, however, only a single clone of the complete *V. anglica* *ssrRNA* gene was sub-cloned and sequenced. Examination of multiple clones would have been expensive and time consuming for the relatively small amount of information additional clones would have yielded. To identify errors, which may have been introduced during amplification or DNA sequencing, the primary and secondary structure of the *V. anglica* *ssrRNA* was compared to other eukaryotic rRNAs.

Examination of potential non-canonical pairs identified six pairs which may result from a misincorporated nucleotide. Possible errors are apparent in the consecutive G:A and A:A pairs in helix 4. Primary and secondary sequence analysis suggest that the adenine bases in positions 34 and 35 are usually occupied by uracil and cytosine residues which would provide the complementary base-pairing. Similarly, sequence analysis suggests that the uracil at position 618 of the U:U pair is probably an adenine. The G:A pair in helix 49 is usually a canonical base-pair, however, sequence variation at positions 1814 and 1917 does not indicate which base if any is likely to be incorrect. The presence of the non-canonical U:U and C:A pairs in helix 50 is also abnormal, however in this instance the non-canonical base-pairing is probably a result of the adenine at position 1953 and the uracil at position 1956 being introduced as part of the reverse amplification primer. Considering that *Taq* DNA polymerase extends a mismatched primer-template far less efficiently than a correct primer-template (Gelfand & White, 1990), a non-complementary reverse primer may account for at least some of the difficulty encountered in the amplification of amoeba *ssrRNA* genes in Chapter 2.

Comparisons of primary and secondary structure also identified a further three possible errors. Within conserved unpaired regions, the thymine between helix 30 and 28 at position 1234 is absent in other *ssrRNAs*, and the first adenine in the region between helices 32 and 33 at position 1303 is usually a cytosine. Finally, there is a possible error immediately distal to the bulge in helix E23-7. At this site a non-complementary base-pair is often found, and in some cases an unpaired nucleotide is present. However, in the *V. anglica* *ssrRNA*, the presence of two unpaired nucleotides suggests a complementary nucleotide may be missing from between positions 897 and 898. Further errors may be present within 'helix' E23-1,

but in the absence of sequence and structural conservation in this region, errors are undetectable by comparative methods.

Although, nine possible errors were detected, two of these were probably introduced with the reverse primer, some of the remaining errors may be naturally occurring mutations. From the sequence analysis of multiple clones used in the partial *ssrRNA* sequence analysis of the V7 and V8 regions (Chapter 5), the combined PCR amplification and sequencing error rate was less than 0.25%. If this rate is constant, the complete *V. anglica* *ssrRNA* gene sequence is likely to have up to five errors. The estimation of nine possible errors suggests that some of these are naturally occurring mutations.

### **Development of the *ssrRNA* structure**

Recently, there have been relatively few, and only minor, refinements to the secondary structure model for the eukaryotic *ssrRNA*, despite the dramatic increase in the number of *ssrRNA* sequences now available (Maidak *et al.*, 1996). The proportion of nucleotides involved with *ssrRNA* secondary structure is very similar to the proportion in tRNA (Gutell *et al.*, 1994), the secondary structure of which has been determined by X-ray crystallography (Kim, 1979). The secondary structure of the eukaryotic *ssrRNA* is therefore considered "proven". However, the low proportion of nucleotides involved with higher-order interactions in *ssrRNA* compared to tRNA, suggests that many *ssrRNA* higher-order interactions, particularly those involving elements well separated by secondary structure, have yet to be determined (Gutell *et al.*, 1994).

Current models infer that the secondary and higher-order structure of *ssrRNA* is static, but, the structure of the *ssrRNA* is almost certainly dynamic: translation involves movement - the ribosome must progress along a mRNA molecule to read the codons, and the tRNA moves from one site to another within the ribosome as the amino acid is transferred to the growing polypeptide chain. The rRNAs, which are central to the function of the ribosome, must therefore undergo changes in conformation to accommodate or promote this essential movement. To identify further structural motifs it may be necessary to consider the *ssrRNA* to be a more dynamic structure with regions capable of alternative conformations. Otherwise improvement in our understanding of the structure of the *ssrRNA* may rely on the development of X-ray crystallography for the analysis of ribosomal components.

## **CHAPTER 4**

### **Molecular phylogeny of the amoebae.**

## Introduction

In 1959, the five kingdom system of classification of organisms was proposed. According to this scheme, the kingdom Monera contains all the prokaryotes, and the eukaryotes are represented by the three multicellular kingdoms, the Fungi, Plantae and Animalia, and the unicellular kingdom Protista (Whittaker, 1959; Margulis & Schwartz, 1982). The phylogenetic relationships between the eukaryotic taxa has been an area of controversy. Traditional methods, such as the examination of comparative phenotypes and the fossil record have been used to examine relationships within the multicellular kingdoms, however, these methods have had limited application for examining phylogenetic relationships among the protists. The protists, which include the heterotrophic flagellates, ciliates, amoebae and the autotrophic microalgae, are poorly represented in the fossil record and the extreme phenotypic variation displayed within the group means that character states may be misinterpreted due to parallelism and convergent evolution. Moreover, there has been little agreement on the relative importance of comparative studies of morphology, physiology and biochemistry for phylogenetic inference (Ragan & Chapman, 1978). Despite these short-comings, it is accepted that the first eukaryotes were protists and the multicellular groups evolved from protistan ancestors. However, the specific relationships between protists, fungi, plants and animals are still a matter of debate.

### **Phylogeny of the protists**

Our understanding of protistan phylogeny improved with the advent of electron microscopy (EM) and the examination of ultrastructural features. Enhanced resolving power of EM over light microscopy provided detailed descriptions of the features including the mitochondria, Golgi apparatus, nucleus, microtubules, sub-membraneous structures and the cell surface. The extent of the ultrastructural variation among the protists far exceeded the diversity within and between other eukaryotic kingdoms, and it was soon evident that these features provided far better indicators of phylogenetic relationships than gross morphological characters. Ultrastructural surveys of algae (Dodge, 1973), ciliates (Corliss, 1979), sporozoa, heliozoa (Smith & Patterson, 1986), slime molds (Olive, 1975) and free-living flagellates (Patterson & Larsen, 1991) resulted in new hypotheses about evolutionary relationships and the redefinition of groups (Irvine & John, 1984;

Corliss, 1984; Margulis *et al.*, 1989; Patterson & Larsen, 1991). Despite these ultrastructural studies there remain many uncertainties concerning the phylogeny of protists. For instance, there has been no systematic survey of the amoebae undertaken, and the "diversity of the amoeboid organism remains largely uncharted" (Patterson, 1993).

### **Classification of amoebae**

Historically, amoebae have been classified using morphological characters apparent at the light microscope level, such as the form of their pseudopodia, the possession of tests and other skeletal elements, fruiting bodies and flagellate stages, and the structure and behaviour of the nucleus during mitosis. More recently, ultrastructural studies have provided additional distinguishing features including the form of the mitochondrial cristae, Golgi dictyosomes and the cell surface structure, but in general there are few fine structural features which are useful for phylogenetic inference.

A summary of the classification of the phylum Rhizopoda is presented in Table 4.1. According to this scheme, the phylum is composed of 8 classes (Levine *et al.*, 1980; Page, 1987). The Heterolobosea are naked, commonly lobose, monopodial amoebae with eruptive locomotion, usually with a transitory flagellate stage and sometimes with fruiting bodies. The mitochondrial cristae are flattened and usually discoid and the Golgi apparatus is seldom organised into dictyosomes (Page & Blanton, 1985). The Lobosea are naked or testate amoebae with lobose pseudopodia, non-eruptive movement, without flagellate stages or fruiting bodies. The mitochondrial cristae are tubular and the Golgi apparatus is organised into dictyosomes. The Filosea are naked or testate with hyaline filiform pseudopodia which are often branched. They have no flagellate stages or spores in their life-cycle. The Granuloreticulosea usually have a test with delicate, finely granular or hyaline branching reticulopodia which form an anastomosing network, although some do form a rigid cell coating. The Eumycetozoa are the cellular and plasmodial slime molds. They have filiform subpseudopodia, sometimes transitionally with an apical pair of flagella, and they produce fruiting bodies usually with a stalk tube. The Caryoblastea represented by a single genus *Pelomyxa* are naked, multinucleate, thickly cylindrical or ovoid with sparse non-motile flagellum-like structures on the cell surface. They have no Golgi dictyosomes or mitochondria, although bacterial symbionts are present. The Plasmodiophorea are intracellular parasites with minute plasmodia. Zoospores



**Table 4.1** Classification of the amoebae (Levine *et al.*, 1980; Page, 1987). The genera shown in bold are those for which ssrRNA sequence data is available, including *Vannella*, the focus of this study.

## Phylum Rhizopoda Von Siebold, 1845

### Class Heterolobosea

- Order Schizopyrenida
  - Family Vahlkampfiidae
    - Valkampfia*, *Naegleria*,  
*Tetramitus*, *Paratetramitus***
  - Family Gruberellidae
- Order Acrasida
  - Family Acrasidae
  - Family Gutulinopsidae

### Class Lobosea

- Subclass Gymnamoebia
  - Order Euamoebida
    - Family Amoebidae
    - Family Thecamoebidae
    - Family Hartmannellidae
      - Hartmannella***
    - Family Vannellidae
      - Vannella***
    - Family Paramoebidae
    - Family Vexilliferidae
  - Order Leptomyxida
    - Suborder Rhizoflabellina
      - Family Flabelluidae
      - Family Leptomyxidae
    - Suborder Leptoramosina
      - Family Stereomyxidae
      - Family Gephyramoebidae
  - Order Acanthopodida
    - Family Acanthamoebidae
      - Acanthamoeba***
  - Order Loboreticulatida
    - Family Corallomyxidae
- Subclass Testacealobosia
  - Order Himatismenida
    - Family Cochliopodiidae
  - Order Arcellinida
  - Order Trichosida

### *Incertae sedis* within subclass Gymnamoebia

- Family Echinamoebidae
- Family Entamoebidae
  - Entamoeba***
- Family Hyalodiscidae

**Table 4.1** continued

**Class Filosea**

Subclass Aconchulina

Order Crisidiscodida

Family Nucleariidae

Family Pompholyxophryidae

Order Cristivesiculatida

Family Vampyrellidae

Family Arachnulidae

Subclass Testaceaiflosia

*Euglypha, Paulinella*

**Class Granuloreticulosea**

Order Athalamida

Family Biomyxidae

Order Promycetozoida

Family Reticulomyxidae

Order Monothalamida

Order Foraminiferida

**Class Eumycetozoa**

Subclass Protosteliia

Order Protosteliida

Subclass Dictyosteliia

Order Dictyosteliida

*Dictyostelium*

Subclass Myxogastria

Order Echinosteliida

Order Liceida

Order Trichiida

Order Stemonitida

Order Physarida

*Physarum*

**Class Caryoblastea**

Order Pelobiontida

Family Pelomyxidae

**Class Plasmodiophorea**

Order Plasmodiophorida

**Class Xenophycophorea**

Order Psamminida

Order Stannodida

***Incertae sedis* within Phylum Rhizopoda**

Family Phreatamoebidae

*Phreatamoeba*

produced in zoosporangia bear an anterior pair of flagella. Sexuality has been reported in some species. Amoebae of the final class, the Xenophyophorea, have a multinucleate plasmodium enclosed in a branched-tube system composed of organic substance with numerous barite crystals and distinct dark faecal pellets (Sleigh, 1989).

One amoeba genus, *Phreatamoeba*, cannot be placed within any existing class due to its unusual combination of characters. In common with the Caryoblastea, *Phreatamoeba* is naked, usually multinucleate, without mitochondria or Golgi dictyosomes, although bacterial symbionts are absent. *Phreatamoeba* also has a flagellate stage and centrioles, and the nuclear envelope does not disintegrate during mitosis.

The Rhizopoda are clearly a very diverse group of amoeboid organisms, and the relationships between the different forms are poorly understood, far more so than for any other group of protists. It is conceded that the Rhizopoda are polyphyletic, and the amoeboid form has evolved several times throughout the evolution of the eukaryotes. In an attempt to reduce the polyphyly of the phylum, alternative classification schemes have been proposed (Cavalier-Smith, 1981, 1986; Margulis & Schwartz, 1982; Lee *et al.*, 1985; Margulis *et al.*, 1989).

In general, classification schemes have removed six classes from the phylum Rhizopoda, leaving just two; the Lobosea and Filosea (Schuster, 1989). A majority of the classes excluded have themselves been increased to the rank of phylum (Margulis & Schwartz, 1982; Corliss, 1984; Margulis *et al.*, 1989). Within the class Eumycetozoa, both the cellular and the plasmodial slime molds have been elevated to the rank of phylum (Martin *et al.*, 1983; Margulis *et al.*, 1989).

In a more recent classification system, several high level taxonomic ranks have been introduced to accommodate the vast diversity of the protists, and immense importance has been apportioned to the presence or absence of Golgi dictyosomes and when present, the type of mitochondrial cristae (Cavalier-Smith, 1993). Due to the absence of Golgi dictyosomes and mitochondria, the unusual *Phreatamoeba* and *Pelomyxa* have been placed within the kingdom Archezoa, which also includes the diplomonads and the microsporidians. Within the kingdom Protozoa, the Heterolobosea are placed in the subkingdom Adictyozoa, due to the absence of Golgi

dictyosomes and the presence of discoid mitochondrial cristae. All the remaining amoeboid taxa are found within various phyla in the subkingdom Dictyozoa due to the presence of dictyosomes. Moreover, the naked anaerobic amoeba *Entamoeba* has been moved from the subclass Gymnamoebia, class Lobosea to a newly erected 'parvkingdom' Entamoebia.

The shift away from the grouping of amoeboid organisms in a single phylum and the upgrading of taxa to higher ranks is clearly a positive move to acknowledge the diversity of the amoebae, and an attempt to reduce the polyphyly within amoeboid taxa. However, the radical re-organisation of protist classification based on essentially two characteristics and the erection of several new ranks, and nomenclature revisions introduced as a consequence, are controversial. Although the diversity of the amoebae is now being recognised, the lack of useful phylogenetic features has meant that our understanding of the phylogenetic relationships among the amoebae lags well behind other protist groups, and the evolutionary relationships between amoebae and other eukaryotic organisms remains unclear.

Recently, the use of molecular data has provided fresh impetus for the examination of phylogenies and eukaryotic evolution, and for the development of a natural classification system based on phylogenetic relationships. Unlike phenotypic and ultrastructural character data, molecular data provides a quantitative measure of relatedness between organisms allowing phylogenetic relationships to be examined from many diverse organisms simultaneously.

### **Molecular data**

In 1965, it was demonstrated that quantitative data can be derived by comparison of macromolecular sequences and this could be used to infer phylogeny (Zuckerkandl & Pauling, 1965). In the first instance, phylogenies were inferred using protein sequences from cytochrome C (Fitch & Margoliash, 1967), ferredoxins (Yasunobu & Tanaka, 1975) and superoxide dismutases (Lumsden & Hall, 1975), followed by actin proteins (Loomis & Smith, 1990) and elongation factors (Hasegawa *et al.*, 1993). The use of protein sequences, however, has been largely replaced by ribosomal RNA sequences for estimating phylogenies. Initially, the 5S rRNA (Hori & Osawa, 1986) was used but due to the insufficient number of phylogenetically informative characters this rRNA was superseded by the large subunit rRNA (De Rijk, *et al.*, 1995) and the small-subunit rRNA (Sogin, 1989; Embley *et al.*, 1994).

## **Ribosomal RNA sequences**

The ribosomal RNA molecules are particularly attractive markers for phylogenetic analyses. Unlike proteins, there is no danger of comparing paralogous genes and they do not undergo lateral gene transfer (Sogin & Elwood, 1986), and the variation in the repeated sequences is very low (Dover, 1982; Schlegel *et al.*, 1991). The rRNAs are universally distributed in all organisms, as well as in mitochondria and chloroplasts. They possess highly conserved regions which can be used to examine relationships between distant taxa, and more variable regions which provide information on the relatedness of similar taxa. The molecule varies in size from 1200 to 2300 nucleotide bases providing in excess of 1000 independent variable characters for analysis. Moreover, the presence of highly conserved regions at the termini of the gene allow amplification of the whole gene using PCR and rapid DNA sequencing (Elwood *et al.*, 1985; Medlin *et al.*, 1988). The small-subunit rRNA has become the most widely used phylogenetic character. The current ssrRNA sequence database has in excess of 2200 complete prokaryotic ssrRNA sequences (including mitochondrial and chloroplast) and 440 eukaryotic sequences and a series of auxiliary programs which can be used for phylogenetic analysis (Maidak *et al.*, 1996).

## **Molecular phylogeny of the prokaryotes**

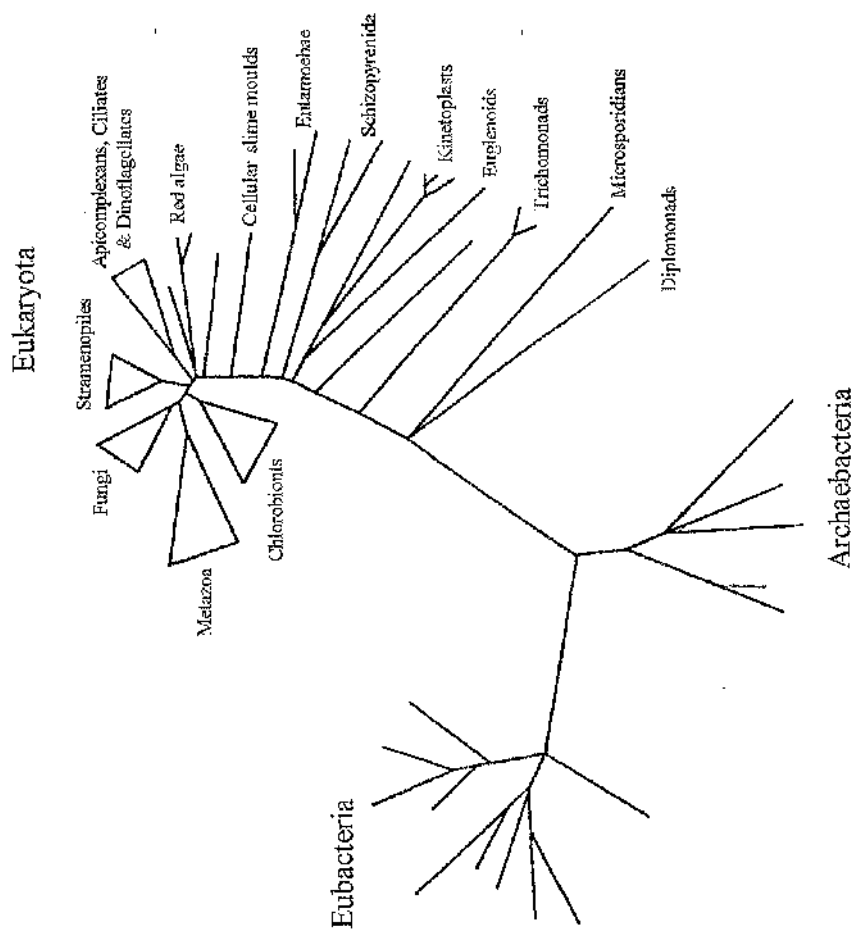
The use of the ssrRNA molecular marker has had an important impact on our understanding of the diversity and phylogenetic relationships of the prokaryotes. Comparisons of ssrRNA prokaryotic sequences has identified a dichotomy within the prokaryotes. The Archaeobacteria lineage, composed of the extreme thermophiles, extreme halophiles, methanogens and some sulphur-dependent bacteria, has been segregated from the Eubacteria lineage which comprises of Gram-positive bacteria, green non-sulphur bacteria, cyanobacteria, proteobacteria, flavobacteria and some others (Woese *et al.*, 1978; Woese, 1987). Further divisions of the Archaeobacteria (Woese *et al.*, 1990; Lake, 1987b) and the pursuit of the root of the primary divisions of life, and thereby the universal ancestor (Sogin, 1991; Pühler *et al.*, 1989; Iwabe *et al.*, 1991), remain areas of lively debate. It is certain, however, that the extent of prokaryotic diversity will continue to increase as more unique lineages are discovered as unculturable organisms are sampled from different environments (De Long, 1992; Fuhrman *et al.*, 1992).

## Molecular phylogeny of the eukaryotes

Despite the revelations of cytological examinations, the extent of eukaryotic diversity was not fully appreciated until *ssrRNA* sequences, from a variety of eukaryotes, were examined. Comparisons of eukaryotic *ssrRNAs* have demonstrated that eukaryotes have evolved over a long period during which many protist lineages have diverged before the relatively recent evolution of the plants, animals and fungi. The diversity of the protists is immense, exceeding the diversity of not only the other eukaryotic kingdoms, but also the Eubacteria and Archaeobacteria (Sogin, 1991; Schlegel, 1994b). The molecular phylogeny of the eukaryotes, determined using *ssrRNA* sequence comparisons, and the relative divergence of the Eubacteria and Archaeobacteria lineages are shown in Figure 4.1.

Although there is a large database for eukaryotic *ssrRNA* genes, it has to be emphasized that most information is biased in favour of certain groups such as the metazoa, fungi and ciliates. The majority of the ancient eukaryotic groups are poorly represented, and data from many of the cryptomonads, haptophytes and most of the amoebae are still lacking. The analysis of *ssrRNA* genes provides a unique opportunity to examine the phylogeny of the amoeboid organisms. However, there is currently complete *ssrRNA* sequence data available for only 12 genera spanning four of the original eight classes of Rhizopoda (Table 4.1).

The first amoeba *ssrRNA* sequence was determined from the cellular slime mold *Dictyostelium discoideum* (McCarroll *et al.*, 1983). At that time, this taxon represented the first eukaryotic branch from the *ssrRNA* phylogenetic tree. Since then several eukaryotic taxa have been shown to branch before *D. discoideum*. The first taxa to diverge at the base of the eukaryotic tree are the anaerobic diplomonads, microsporidians and trichomonads (Sogin, 1989). The branching order of these three taxa has remained an area of debate and is still unresolved. Initially, the examination of the *ssrRNA* from *Vairimorpha necatrix* identified the Microsporidia as the most ancient eukaryotic organisms (Vossbrinck *et al.*, 1987). This is supported by their prokaryotic-like sized ribosomes and rRNAs, the lack of centrioles, dictyosomes, kinetosomes, flagella and a microtubular cytoskeleton. Since then, analysis of the diplomonad *Giardia lamblia* suggested that the diplomonads, which possess flagella, branched prior to the microsporidians (Sogin *et al.*, 1989). However, the large variation in the guanine and cytosine (G+C) composition of these organisms meant that the branching order was influenced by outlying prokaryotic



**Figure 4.1** Molecular phylogeny of the eukaryotes (Eukaryota). The lengths of the branches correspond to genetic distances. Redrawn from Schlegel (1994b).

taxa with different G+C contents. The diplomonad *Hexamita inflata*, which has a G+C content more characteristic of eukaryotic ssrRNAs, was used to overcome the uncertainty due to G+C bias. Despite this, the branching order at the base of the eukaryotic tree remains unclear. Recent phylogenies have demonstrated that the trichomonads, which possess kinetosomes and flagella, diverged before the diplomonads (Leipe *et al.*, 1993). Examination of additional diplomonad sequences may complement partial microsporidian sequences (Zhu *et al.*, 1994; Vossbrinck *et al.*, 1993) and complete trichomonad sequences (Gunderson *et al.*, 1995), to resolve the branching order of these early eukaryotic taxa in the near future.

After the acquisition of mitochondria by the uptake of a bacterial endosymbiont (Yang *et al.*, 1985), the next taxon to branch from the eukaryotic tree is the euglenoids (Sogin *et al.*, 1986a). These are a group of flagellates comprising of the heterotrophic kinetoplasts, which include the free-living bodonids and the parasitic trypanosomes, and the free-living euglenids.

After this point several amoeboid lineages branch from the eukaryotic tree. The first lineage is the Schizopyrenida which was first represented by *Naegleria gruberi*. The phylogenetic position of this amoeba-flagellate provided the first molecular evidence to support the polyphyletic origin of amoebae (Clarke & Cross, 1988). Small-subunit sequence comparisons demonstrated that *N. gruberi* was not related to the naked, lobose amoeba, *Acanthamoeba castellanii*, which branched much later in the eukaryotic tree. The molecular evidence for the polyphyly of the amoebae strongly supported mitotic and ultrastructural features. The order Schizopyrenida was classified within the class Lobosea, phylum Rhizopoda (Page, 1976), but the persistence of the nucleolus and nuclear membrane through mitosis, along with the discoid mitochondrial cristae and the absence of Golgi dictyosomes demonstrated that the schizopyrenids were not related to the Lobosea, of which *A. castellanii* is a member. As a consequence, a new class, Heterolobosea, was erected to accommodate Schizopyrenida (Page & Blanton, 1985), and since that time the Heterolobosea have been moved to a new phylum to respect these significant ultrastructural and genetic differences (Cavalier-Smith, 1993). Schizopyrenida ssrRNA sequences from the genera *Vahlkampfia*, *Paratetramitus* and *Tetramitus* (Hinkle & Sogin, 1993), support the phylogenetic position of the Schizopyrenida and the probable flagellate ancestry of the Heterolobosea (Clark & Cross, 1988).



The next amoeboid lineages are the plasmodial slime mold represented by *Physarum polycephalum* (Johansen *et al.*, 1988), which is not shown in Figure 4.1, and the parasitic anaerobic amoeba genus *Entamoeba histolytica* (Sogin, 1989). These are followed by the cellular slime mold, *D. discoideum*, and the free-living anaerobic amoeba *Phreatamoeba balamuthi* (Hinkle *et al.*, 1994), which is also not shown in Figure 4.1.

The absence of both mitochondria and Golgi dictyosomes in *P. balamuthi* may reflect a more primitive ancestry than the phylogenetic position suggests, unless the loss of both of these features has resulted from reduction as an adaptation to a micro-aerobic habitat. The presence of a flagellate stage and centrioles during mitosis also clearly distinguish *Phreatamoeba* from *Entamoeba*, demonstrating that these anaerobic amoebae represent separate lineages (Chávez *et al.*, 1986).

The plasmodial slime molds and cellular slime molds, which have dictyosomes, mitochondria with tubular cristae and complicated life-cycles with fruiting bodies, were placed in the class Eumycetes (Page, 1987). However, these organisms are quite different with respect to their life cycles and cellular differentiation and are now placed in separate phyla as they may not be related by a common ancestry (Margulis *et al.*, 1989). The molecular evidence supports this hypothesis as their phylogenetic positions on the eukaryotic tree are separated by the *Entamoeba* lineage. Despite the apparent position of *Phreatamoeba*, the cellular slime mold *D. discoideum* is usually recognised as the last taxon to branch from the eukaryotic tree before the divergence of the crown taxa (Embley *et al.*, 1994).

### **Molecular phylogeny of the crown taxa**

Approximately one billion years ago eukaryotic evolution went through a stage of rapid diversification which led to a series of lineages including the red algae, the apicomplexa, ciliates and dinoflagellates, the stramenopiles, the fungi, the metazoa and the chlorobionts (Sogin, 1991). Together, these relatively recent diverging taxa are known as the 'crown' taxa (Knoll, 1992). The branching order of the lineages among the crown taxa is difficult to determine reliably because they are only separated by small amounts of sequence differences (Embley *et al.*, 1994). Sequence analyses has demonstrated, however, that the separate lineages are quite robust and largely reflect monophyletic groupings.

Despite considerable variation in the ultrastructure and ecology of the dinoflagellates, apicomplexa and ciliates, all these organisms possess sub-membranous alveolar structures (Taylor, 1976). The significance of alveoli was not realised until ssrRNA sequence data demonstrated that the apicomplexa and dinoflagellates are sister-groups of the primarily phagotrophic ciliates (Gajadhar *et al.*, 1991). Collectively, due to the common alveolar structure, members of this monophyletic group are known as the alveolates (Patterson & Sogin, 1993). Sequence data also supports the stramenopile clade (Leipe *et al.*, 1994), a disparate group consisting of the brown algae and the diatoms, chrysomonads, water molds, opalinids and others whose relationship was suggested on the basis of the presence of tripartite hairs (Patterson, 1989). The monophyly of the metazoa, fungi and chlorobionts (green algae and plants) lineages are also supported by ssrRNA sequence analyses (Van de Peer *et al.*, 1992; Wainright *et al.*, 1993; Wilcox *et al.*, 1993).

Within the crown taxa two amoeboid lineages have been identified, although neither is shown in Figure 4.1. The first lineage was determined from sequence analysis from the naked, lobose amoeba *A. castellanii* of the subclass Gymnamoebia, class Lobosea (Gunderson *et al.*, 1986). This amoeba branches within the crown taxa close to the plants and fungi (Sogin *et al.*, 1986b; Wainright *et al.*, 1993). A second amoeba from this subclass, *Hartmannella vermiformis*, has since been sequenced (Gunderson *et al.*, 1994; Weekers *et al.*, 1994) and these two gymnamoebae form a clade (Weekers *et al.*, 1994). The second lineage was identified more recently after examination of two filose, testate amoebae, *Euglypha rotunda* and *Paulinella chromatophora*, from the class Filosca (Bhattacharya *et al.*, 1995). These filose testate amoebae form a robust clade with the Chlorarchniophyta, an independent lineage of photosynthetic alga with reticulate pseudopodia (Grell, 1990) which are thought to be derived from an amoeba which acquired a chloroplast by retaining part of an engulfed eukaryotic algae (McFadden *et al.*, 1994a). This clade falls within the crown of the eukaryotic tree although there is no support that this amoeboid lineage is a sister group with any particular crown taxon (Bhattacharya *et al.*, 1995).

From the 12 genera of amoebae so far sequenced, seven separate amoeboid lineages have been identified. However, considering that several diverse classes of amoebae have not been examined, it is highly probable that further amoeboid lineages will

be identified as more sequence data becomes available. In the absence of useful morphological features, it is clear that sequence information can play a key role in the development of our understanding of the phylogeny of amoebae and the evolution of eukaryotic organisms.

### **Phylogenetic analysis of rRNA sequence data**

The examination of ssrRNA sequence data for phylogenetic purposes is far from a simple process. The first obstacle to overcome when using sequence data is the choice of characters to be used in the analysis. For all phylogenetic analyses it is essential to compare equivalent nucleotide positions. To accomplish this, the nucleotide sequences are aligned using primary and secondary structure and those nucleotide positions which cannot be unambiguously aligned are excluded from the analysis. Since there are no rigid rules governing alignment or to justify the inclusion or exclusion of characters within the alignment, this fundamental aspect of data handling for phylogenetic analysis is open to criticism, but is ultimately left to the discretion of the researcher.

If all the nucleotide differences between aligned sequences (ie. character state changes) were truly indicative of evolutionary relationships there would be little difficulty in reconstructing phylogenies. However, character states may arise which produce misleading phylogenetic information. These character state transformations, collectively known as homoplasies, may occur from either parallel or convergent evolution or from the reversal of a character state from the advanced to its ancestral condition. Homoplasies are a particular problem if rapidly evolving sites are included or there is significant variation in the rate of evolution between different lineages, because repeated character state transitions may occur at the same nucleotide position, disguising the true number of evolutionary changes. Misleading phylogenies may also occur if the inference method assumes that the rate of evolution is constant for each nucleotide position, because some regions evolve slowly as the sequence is under functional constraint, whereas unconstrained regions may mutate more readily (Yang, 1993). Furthermore, some character state changes occur more frequently than others; transition substitutions (substitutions between two purine bases or between two pyrimidines) are more frequent than transversion substitutions (substitutions between pyrimidines and purines) (Brown *et al.*, 1982). Phylogenetic inference methods also assume that all characters are independent with respect to change. This fundamental assumption

is violated because compensatory mutations are more likely to occur in stem positions to maintain secondary structure. However, stem positions are phylogenetically informative, and their removal reduces the chance of finding the correct tree (Dixon & Hillis, 1993). Finally, phylogenies may also be sensitive to specific taxa, too many taxa, the order in which taxa are included in the analysis or bias in the base frequencies (Sogin *et al.*, 1989; Leipe *et al.*, 1993; Wainright *et al.*, 1993). The problems associated with inferring phylogenies from sequence data are now relatively well understood and the majority of phylogenetic inference methods now include evolutionary models or other measures which take into account the sources of error described.

### **Phylogenetic inference**

Phylogenies have been inferred from rRNA sequence data principally by three methods; parsimony and maximum likelihood methods which infer phylogenies directly from the character data, and distance data matrix methods. The choice of method will depend on a variety of factors including the composition of the data set, the degree of homoplasy, the computation time, which evolutionary models are used and if statistical analysis is required on the phylogenetic tree (Swofford *et al.*, 1996).

### **Parsimony methods**

Under the principles of parsimony, estimates of phylogenetic trees are derived by finding the "simplest" explanation for the character states which have been inherited by taxa from a common ancestor. For nucleotide sequence data, this equates to the minimal number of evolutionary steps (nucleotide substitutions) required to explain a given data set. Because of its simplicity, parsimony methods do not require a specific evolutionary model, although it does assume that evolutionary change is rare. If this assumption is significantly violated, multiple base substitutions will go undetected and parsimony analysis may fail. Parsimonious trees are evaluated by calculating the tree length i.e. the total number of base substitutions required to produce the tree. The most parsimonious tree has the lowest tree length.

As pointed out above, the simplicity of parsimony means that an evolutionary model is not required, although generalised parsimony programs are now available which allow weights to be assigned to different character state transformations (Fitch, 1971; Sankoff, 1975). Alternatives to the generalised method have been developed

to reduce the amount of homoplastic data used in the analysis. Transversion parsimony reduces homoplasy by completely ignoring all transition substitutions, only including transversion substitutions in the analysis. Evolutionary parsimony (Lake, 1987a), makes a distinction between transversion and transition substitutions and uses more data than transversion parsimony, but is restricted to sets of four taxa and has produced controversial results.

At the beginning of a parsimony search an initial tree is obtained and the length of the tree is determined using the parsimony algorithm. The tree is then rearranged to find the shortest, most parsimonious tree. There are three basic tree searching methods which seek to find the shortest tree. An exhaustive search examines all possible trees, but is very time consuming. Branch and bound methods take shortcuts to find the optimal tree, but are again very time consuming and may only find the most parsimonious tree for 20 to 25 taxa in a realistic time scale (Hendy & Penny, 1982). For a quicker estimation of the tree, or for larger numbers of taxa, a variety of heuristic tree search algorithms are available, although they may not necessarily find the most parsimonious tree.

### **Maximum likelihood**

A maximum likelihood approach to phylogenetic inference evaluates the probability that a chosen evolutionary model will generate the observed data set. The first use of maximum likelihoods for phylogenetic inference was with gene frequency data (Edwards & Cavalli-Sforza, 1964), but the methodology has since been adapted for nucleotide sequence data (Felsenstein, 1981, 1993). There are a variety of evolutionary models which are used for maximum likelihood analysis, although they are all intrinsically related (Swofford *et al.*, 1996). A majority of the models do not assume equal rates of change throughout the tree or equal base frequencies, and they allow for differences between transition and transversion rates to be accounted for.

Under the assumption that nucleotide sites evolve independently, the likelihood for each nucleotide position is calculated separately. The product of the individual site likelihoods is computed to determine the joint probability for the tree under the given model. As the probability of each single observation is a very small number, the log of the likelihood for each site is evaluated and summed for the log likelihood, which is used to describe the tree.

Maximum likelihood is useful because it is least affected by sampling error and therefore tends to be robust to many violations of the assumptions used in its models, although maximum likelihood analyses are computationally expensive with a single run requiring many hours.

### **Pair-wise distance methods**

Until recently, computers were too slow and algorithms too inefficient to exploit maximum likelihood techniques. Distance methods played a more important role in determining phylogenies from sequence data. With these methods, distances or branch lengths are computed between each pair of taxa from nucleotide sequences using an evolutionary model to correct for multiple base substitutions and variable base frequencies. The phylogenetic tree may then be determined using various of methods including cluster analysis (Sneath & Sokal, 1973), Distance Wagner and its variants (Farris, 1972; Swofford, 1981), Fitch-Margoliash (Fitch & Margoliash, 1967) or neighbour joining (Saitou & Nei, 1987). Although analyses using distance data are very fast, the conversion of character data to distance measures involves discarding an enormous amount of potentially useful data, and there is no immediate information available with regard to the support for the tree.

### **Bootstrap resampling**

To provide an indication of the confidence of the branching order of a phylogenetic tree the character data set can be resampled using the bootstrap technique (Felsenstein, 1985). With this method the data set is resampled in excess of 50 times. Each resampled data set is obtained by randomly choosing characters (nucleotide positions), with replacement, up to the total number of characters in the original data set. The variation among the resampled data sets is then taken to indicate the size of the error involved in making estimates from the original data set. The percentage number of times that a clade is present in the trees from the resampled data sets provides an indication of the confidence for each clade in the tree. The bootstrap analysis is the best known method for assessing the stability of phylogenetic trees. Generally, bootstrap analysis provides a conservative measure of the confidence of tree topologies and computer simulations have shown that values as low as 70% can define known correct topologies (Wainwright *et al.*, 1993).

## Summary of chapter

In this chapter, the *ssrRNA* gene from the naked non-sporulating fan-shaped amoeba *Vannella anglica* (Rhizopoda, Lobosea, Gymnamoebia, Euamoebida, Vannellidae) was used to determine its phylogenetic position in the eukaryotic tree, and to examine potential relationships with amoebae and other organisms. The entire *ssrRNA* gene sequence of *V. anglica* (Chapter 3) was aligned with several eukaryotic *ssrRNA* sequences using both primary and secondary structure information. Phylogenetic analysis was carried out using homologous nucleotide sites with parsimony and maximum likelihood phylogeny inference programs.

## Materials and methods

### Sequences

To examine the phylogenetic position of *Vannella anglica*, the *V. anglica* sequence from Chapter 3 was compared with ssrRNA sequences from a variety of eukaryotic taxa. Unaligned ssrRNA sequences from two filose testate amoebae, *Paulinella chromatophora* and *Euglypha rotunda*, were obtained from the Genetics Computer Group GenEMBL database and in excess of 40 pre-aligned ssrRNA sequences were obtained from the Ribosomal Database Project (Maidak, *et al.*, 1996). A listing of the ssrRNA sequences used in the analysis, including the short identification, GenEMBL accession numbers and references are given in Table 4.2.

### Alignments

All aligned sequences were trimmed to 4017 nucleotides to remove characters present at the termini of some sequences which are not recognised by phylogenetic programs. The *V. anglica*, *P. chromatophora* and *E. rotunda* ssrRNA sequences were initially aligned with the most similar aligned eukaryotic ssrRNA sequences using the ALIGN\_SEQUENCE program (Maidak *et al.*, 1996). The three aligned sequences were then down-loaded and analysed with up to 20 other aligned sequences with the ESEE sequence editor. Minor improvements to the alignments were made when the three sequences were compared to larger numbers of sequences. Final improvements to the alignment were made when the secondary structure of the *V. anglica* ssrRNA (Chapter 3) was considered. The alignment of the nucleotide positions in each helix was checked with several sequences for which secondary structure information is available. Alignment of sequences was relatively easy over the highly conserved regions, but in hyper-variable regions the presence of expansion segments and quickly evolving sequence makes unambiguous alignment of homologous nucleotide positions impossible. The order of the aligned sequences used in the input files for the phylogenetic analyses was chosen at random.



**Table 4.2** Alphabetical listing of the eukaryotic ssrRNA sequences.

Short Id	Name of Organism	GenEMBL Accession Number	Reference
Acyl.bisex	<i>Achyla bisexualis</i> *	M32705 J02951	Gunderson <i>et al.</i> , 1987.
Acnc.ungui	<i>Acanthocephalus unguiculata</i> *	L10823	Wainwright <i>et al.</i> , 1993.
Acnt.castl	<i>Acanthamoeba castellanii</i> *	M13435	Gunderson & Sogin, 1986.
Alxn.fundy	<i>Alexandrium fundyense</i> *	U09048	Scholin <i>et al.</i> , 1993.
Babs.bovis	<i>Babesia bovis</i> (strain Samford)*	M87566	Ellis <i>et al.</i> , 1992.
Clims.islan	<i>Chlamys islandica</i> *	L11232	Kenchington <i>et al.</i> , 1993.
Climy.reinh	<i>Chlamydomonas reinhardtii</i> *	M32703 J02949	Gunderson <i>et al.</i> , 1987.
Clra.rptns	<i>Chlorarachnion reptans</i> Grell (Tunisia)*	X70809	McFadden <i>et al.</i> , 1992.
Crcn.nfrmn	<i>Cryptococcus neoformans</i> *	L05427	Fan <i>et al.</i> , 1994.
Crpm.phi	<i>Cryptomonas phi</i> *	X57162	Douglas <i>et al.</i> , 1991.
Crsp.parv3	<i>Cryptosporidium parvum</i> *	L16997	Pieniazek <i>et al.</i> , 1993a.
Dict.disco	<i>Dictyostelium discoideum</i> *	K02641	McCarroll <i>et al.</i> , 1983.
Dros.melan	<i>Drosophila melanogaster</i> *	M21017 M29800	Tautz <i>et al.</i> , 1987.
Emil.huxle	<i>Emiliania huxleyi</i> *	L04957	Bhattacharya <i>et al.</i> , 1992.
Encz.hellm	<i>Encephalitozoon hellem</i> *	L19070	Pieniazek <i>et al.</i> , 1993b.
Entm.histo	<i>Entamoeba histolytica</i> *	X61116	Que & Reed, 1991.
Eugl.rotun	<i>Euglypha rotunda</i> *	X77692	Bhattacharya <i>et al.</i> , 1995.
Fucs.distl	<i>Fucus distichus</i> *	M97959	Bhattacharya <i>et al.</i> , 1992.
Grac.leman	<i>Gracilaria lemaneiformis</i> *	X54263 M54986	Bhattacharya <i>et al.</i> , 1990.
Hart.vrmif	<i>Hartmannella vermiformis</i> *	M95168	Gunderson <i>et al.</i> , 1994.
Hpls.nlsni	<i>Haplosporidium nelsoni</i> *	X74131	Fong <i>et al.</i> , 1993.
Klrm.flecd	<i>Klebsormidium flaccidum</i> *	M95613	Wilcox <i>et al.</i> , 1993.

Table 4.2 continued.

Short Id	Name of Organism	GenEMBL Accession Number	Reference
Lbyr.minut	<i>Labyrinthuloides minuta</i> *	L27634	Leipe <i>et al.</i> , 1994.
Mllm.papil	<i>Mallomonas papillosa</i> *	M55285	Ariztia-Carmona, 1991.
Mus_muscul	<i>Mus musculus</i> *	X00686	Raynal <i>et al.</i> , 1984.
Naeg.grube	<i>Naegleria gruberi</i> *	M18732	Clark & Cross, 1988.
Neur.crass	<i>Neurospora crassa</i> *	X04971	Sogin <i>et al.</i> , 1986b.
Oxyt.granu	<i>Oxytricha granulifera</i> *	X53486	Schlegel <i>et al.</i> , 1991.
Parm.tetra	<i>Paramecium tetraurelia</i> *	X03772	Sogin & Elwood, 1986.
Paul.chrom	<i>Paulinella chromatophora</i> *	X81811	Bhattacharya <i>et al.</i> , 1995.
Phre.balam	<i>Phreatamoeba balamuthi</i>	L23799	Hinkle <i>et al.</i> , 1994.
Phys.pceph	<i>Physarum polycephalum</i> *	X13160	Johansen <i>et al.</i> , 1988.
Plas.falca	<i>Plasmodium falciparum</i> *	M19172	McCutchan <i>et al.</i> , 1988.
Pyrrn.salin	<i>Pyrenomonas salina</i> *	X54276	Eschbach <i>et al.</i> , 1991.
Sacc.cerev	<i>Saccharomyces cerevisiae</i> *	J01353 M27607	Rubtsov <i>et al.</i> , 1980.
Sclt.costa	<i>Skeletonema costatum</i> *	M54988	Medlin <i>et al.</i> , 1988.
Srey.muris	<i>Sarcocystis muris</i> *	M64244 M34846	Gajadhar <i>et al.</i> , 1991.
Sybd.corcu	<i>Symbiodinium corcorum</i> *	M88518	McNally <i>et al.</i> , 1993.
Trtr.foetu	<i>Tritrichomonas foetus</i>	M81842	Chakrabarti <i>et al.</i> , 1992.
Txpl.gondi	<i>Toxoplasma gondii</i> *	X68523	Long <i>et al.</i> , 1993. (Chapter 3).
Vann.angli	<i>Vannella anglica</i> *		Hinkle & Sogin, 1993.
Vlkf.lobos	<i>Vahlkampfia lobospinosa</i>	M98052	Messing <i>et al.</i> , 1984.
Zea_mayW22	<i>Zea mays</i>	K02202	

## Exclusion of nucleotide positions

For phylogenetic analysis of sequence data, it is of paramount importance that only nucleotide positions which are equivalent are included in the analysis. For this reason all hyper-variable sites have been excluded from the phylogenetic analysis. This incorporates helices E8-1, 10, E23.1-E23.5, and E23-9, the termini of helices E10.1, E23.7, 29, 43, 44, 46 and 49, and the E45-1 expansion segment, with respect to the *V. anglica* sequence (Chapter 3), and other extended regions unique to other taxa. Sequences at the start and end of the gene including the primer sequences, and all positions of the *V. anglica* sequence, which were identified as possible polymerase errors in Chapter 3, were also excluded. Out of the 4017 aligned nucleotide positions, 1488 were used for phylogenetic analysis. Positions which contain an ambiguous nucleotide for any one of the taxa were not excluded from the analysis as the phylogeny programs either remove these positions from the analysis themselves or account for the discrepancy.

To show the complete alignment containing all the taxa for each of the 4017 positions would require in excess of 50 pages. A restricted alignment of sequences from five taxa displaying the nucleotide positions in the phylogenetic analysis is given in Figure 4.2. The alignment includes the two pre-aligned ssrRNA sequences *Saccharomyces cerevisiae* and *Acanthamoeba castellanii* (Maidak *et al.*, 1996), and the ssrRNA sequences from *V. anglica*, *P. chromatophora* and *E. rotunda* which have been aligned in this study.

---

**Figure 4.2** Selected eukaryotic ssrRNA sequence alignment. The complete aligned sequence of *V. anglica* is given on the first row. In succeeding sequences, nucleotides which are identical to *V. anglica* are denoted by a dot: only nucleotides which differ are given. Dashes have been used to aid alignment of homologous positions. The 1488 nucleotide positions used in the parsimony and maximum likelihood analyses are marked with an asterisk. The number at the end the top row represents the *V. anglica* nucleotide position of the last base.

**Figure 4.2**

<b>Vann. angli</b>	-----ACCTGGTTGATCCTGCCAGTGAATGCTTGAATTAAAGATTAA-GCCATGCAAAATGT-AGTAT-AAATCA-----	70
Paul. chrom	-A.....G.....AG.C.....TC.C.....TG.C.A.....GCA.C-----	
Eugl. rotun	-A.....A.C.....TC.C.....TG.C.A.....CA.C-----	
Sacc. cerev	-T.T.....AG.C.....TC.C.....TG.C.A.....GCA.-----	
Acnt. castl	-T.....AG.C.....TC.C.....TG.C.A.....GCTTG-----	
<b>Vann. angli</b>	**** * * * * * **** * * * * * **** * * * * * **** * * * * *	
Paul. chrom	-----CTTT-TAC-T--GTGAAAC-TGT-GAAC-GGCTCATTAAT-CAGTATAG-----	113
Eugl. rotun	-T.A.....C.....C-T..A.....T.....CA..A-----	
Sacc. cerev	-T.A.....C.....C-T..A.....T.....CA..TA-----	
Acnt. castl	-T.A.....A.....C-T..T.....C.....C-----	
	-T.A.....G.....C-G-T..C-G-T-----	
<b>Vann. angli</b>	***** ***	
Paul. chrom	TTTATTGATG--GTATTGAGGTTTACCT-----CGAACT-ACAATGG-ATAACTGT-AGTAAC-CTAGAGCT	174
Eugl. rotun	.....T.C-----TT.....C.....TT.....	
Sacc. cerev	.....CC-----TT.....C.....TT.....	
Acnt. castl	.....A.....TCC-----TTT.....T.....C.....TT.....	
	.....C.CTTTTG.C-----TTTTTTT.C.....T.....C.....TT.....	
<b>Vann. angli</b>	*****	
Paul. chrom	AATACATGCA--AAAACAATTCAACTTTT-----	202
Eugl. rotun	.....G-C...A-CC.GG...C.....	
Sacc. cerev	.....-C...G-TCC.....C.....	
Acnt. castl	.....T-T...A-TC.G.CC.....	
	.....G-C.GG-TCC.G.GCGGGGGGCAGGGCT-----TCACGGCCCTGTCTCTGCA	
<b>Vann. angli</b>	*****	
Paul. chrom	-----GGAAGAATTGGCACTTATTAG-ATATCAAAACCAATATATTTTC-----	244
Eugl. rotun	GCG...-CGGGT.T.T.....AA.....CGCG.CC..TG-----	
Sacc. cerev	GCG..G.-GGGAT.T.T.....CA.....CCACC..GG-----	
Acnt. castl	-----G.GAT.T.T.....AA...T...G-TC..C-----	
	TGC.C.GAGGGAT.T.T.....-G.TAA.....GCGGCAGGGGTACG-----	



Figure 4.2 continued.

```

***** ** ***** ***** * ***** * ***** ***** ***** *****
Vann.angli GATGTTA-GTA-TATTGGACTAACA-TGGTG-G-TAACGGG-TA-ACGGGGAAT-TGGGGTTC-GATT-CCGGAGAAGG- 402
Paul.chrom .C.G.-.G-.GA.....C.G-.C-T.....-G-CAC..AG.A-.TA..G.T-CGA.T.....G.-
Eugl.rotun .G.-.G-.AA.....C.....CT-A.....-A.....-.....G.-
Sacc.cerev .G.-.G-.G.....C.....T-T-C.....-AA.....-.....G.-
Acnt.castl .G.-.G-.GA.C.C.C.....C.....A.....-A.....-.....G.-

***** ***** ***** * ***** ***** ***** ***** *****
Vann.angli AGCAT-GAGAAAC-GGC-TACTACT-TCTAAGGA--AAGCAGCA-GGCGCGCAAAATT-ATTG-AATGGA-AACAATTT-- 470
Paul.chrom .C.-.....-C.A-.C.....-G.....-CC.-...CCT-G...C-
Eugl.rotun .C.-.....T.....C.A-.C.....-G.....-CC.-...CCT-G...C-
Sacc.cerev .C.-.....-C.A-.C.....-G.....-CC.-...CCT-.TTC-
Acnt.castl .C.-.....T.....C.....-G.....-CC.-...CCC-G...C-

***** ***** ***** ***** * ***** ***** ***** *****
Vann.angli TCCGA-AATAGTGACAATTAATAACAAT--ACGT-TGCA--TAAATTG---TGA-TGTAATTGGAATGA--- 529
Paul.chrom ACT..-GG...T.GA...C...CG-G.TT---T.T-G---CT-G...G...
Eugl.rotun AGG..-GG...GA...G...CG-G.CT---TCG---CT-G...C...
Sacc.cerev AGG..-GG...A...G...AG-G.CC---ATTGCG---CT-
Acnt.castl GGG..-GG...A...A...AGGC.TC---GA...-GA...CT-

***** ***** ***** ***** * ***** ***** ***** *****
Vann.angli GTACAGTT-TAAACCCT---GTACAGTT-TAAACAATTTAGAG 565
Paul.chrom .GT.A.-.T.....-A-TG...-G.T.C.C.G...
Eugl.rotun .A.-.T.....-G.T.C.C.G...
Sacc.cerev .A.G.-.TA...-G...G...
Acnt.castl .A.-.C.....-G...G...

** *** ***** ** ***** ***** ***** ***** ***** *****
Vann.angli GG-CAA-G-TC-TGGT-GCCA-GC-AGCCGGG-TAATTCAGCTCTAATAGTGTATATCTATG-TTGT--GCAGTTA- 633
Paul.chrom .-.-.-.-.-C.G...C...TA.AA-
Eugl.rotun .-.-.-.-.-C.G...C...TA.A-
Sacc.cerev .-.-.-.-.-C...C...TA.A-
Acnt.castl .-.-.-.-.-C...C...TA.A-

```

Figure 4.2 continued.

```

*****
Vann. angli -AAAGCTCGTAGTT-GAATTTTG--AATA-TTTAAA-TATGTTT- 672
Paul.chrom -...N.....-G...CT--GG-T-GAC-CG-GG-CACG-TCG-
Eugl.rotun -.....-G...CT--G..G-AAC-C..A.-T.GG-CCC-
Sacc.cerev -.....C.....-GGC-CCG-GT-.GGCCGG-TCC-
Acnt.castl -.....-G..C.A.-GGAC-GC--GC-AT-T.-
-----
Vann. angli -----
Paul.chrom -----
Eugl.rotun -----
Sacc.cerev -----
Acnt.castl -----
-----CAAGCGCCCGTGTGTCGGGTCAAACGGCGACTGCGTTGGCGTTGCGGGCTCGGTC
-----
Vann. angli -----GGTTACTTTTATAAGTATGACTACTTATTATATCCTTCCCTACTAG--TAACTCGATCTGTTCTTAAATTG 741
Paul.chrom -----G..TC..T--CGGGC--
Eugl.rotun -----GTTC..T--TCGGG.-
Sacc.cerev -----GAT--T..TCG--
Acnt.castl CGTCGGTG.ACCCT.G.GGTCTTAATC.G.GTG.C.ACGGCC.GCC.G.C.CCTCCT.CTGGATT.C.-GT.CCTG--
-----
Vann. angli AATGGAATAGCATCTTTTCGGCGTGTGCGATAGCATGTATATCTCCTCCAATAAGTGGAGAGCGGT--TAAACCTT-AA 817
Paul.chrom -----CGT.TG.GCG.-----CCT.C.-TTT.C-----C.TCCTTCCAG.
Eugl.rotun -----ATGAGTCCGC.-----TG.GT.-T.T-----C.TCCTTCCAG.
Sacc.cerev -----GTAC.GGA.TT-----CCA.C.G.-GCCTTT-----CCTTCTGGC.-..
Acnt.castl -----CTATTGAG.-----GTG.G.-.CGTCA-CAGGGGCT.-A.CGT
-----
*****
Vann. angli AG-----TTTATTG-AGTGGATCAAGATATATATTTAATTGGAG-TTC-----TAGTTCATTACTTGTG-AGAAAATT 884
Paul.chrom GAG-TTGTCCTACC-----GCTCGTTCGGCA-GG-GCGCT.C.-A.-----G.A..T.....-A.....
Eugl.rotun GAG-CTGCG.CTACC-----T.T.CTCGGCCG-GA-CG.AT.C.-A.-----G.A..TA.....-A.....
Sacc.cerev CC-T-..GAG.CC-----TT-GT-GGC-----C-TTG-.CGA-AC.-----GA.T.....-A.....
Acnt.castl C.T-CATGCAA..G.-----CG-GCG.TGG-----G.-----CCCT.G.-GC.-----C..A..G.....CG.....-A.....

```

Figure 4.2 continued.

*****			*****		
Vann.angli	CAGTGTTTAAGGC-AACCTCITT	-----GGGGTTC-TGA-ATACAT-TAGCA	926		
Paul.chrom	.....CC.A...G.	-----TTAC.C.T-			
Eugl.rotun	.....A...GG.	-----TTAT.C.G-			
Sacc.cerev	.....C.A...GG.	-----GTATT.C.-C...T.			
Acnt.castl	.....C.A...GG.AGA.CCAATT	-----TTCT.CCAC-C.			
*****			****		
Vann.angli	TGGGATAATAAAAT	-----AGGAT--TT-AAAT	951		
Paul.chrom	...A.....C.	.....C-.-CTG.			
Eugl.rotun	...A.....TT.	.....C-.-TGG.			
Sacc.cerev	...A.....G.	.....C-G.-TGG.			
Acnt.castl	.....GG.	.....CCC.G-TCC.			
Vann.angli	-----	-----			
Paul.chrom	-----	-----			
Eugl.rotun	-----	-----			
Sacc.cerev	-----	-----			
Acnt.castl	-----	-----			
Vann.angli	-----	-----			
Paul.chrom	-----	-----			
Eugl.rotun	-----	-----			
Sacc.cerev	-----	-----			
Acnt.castl	-----	-----			
Vann.angli	-----	-----			
Paul.chrom	-----	-----			
Eugl.rotun	-----	-----			
Sacc.cerev	-----	-----			
Acnt.castl	-----	-----			



Figure 4.2 continued.

```

*****
Vann.angli  TATATATT-GTTGGTTT-----GAATAAT--T 976
Paul.chrom  C..T.....CT.G.GC--C
Eugl.rotun  C..T.....CT.GG.C--
Sacc.cerev  C..T.....CT.GG.C--C
Acnt.castl  CC..T..CA.....TGGCAGCGG.CG.C--

***
Vann.angli  TAAATAA----- 983
Paul.chrom  G..G.....
Eugl.rotun  G..G.....
Sacc.cerev  ATCG.....
Acnt.castl  AGGG.....

Vann.angli  -----
Paul.chrom  -----
Eugl.rotun  -----
Sacc.cerev  -----
Acnt.castl  -----

*****
Vann.angli  TGATTAATAGGGATAG-TT- 1001
Paul.chrom  .....G.....
Eugl.rotun  .....G.....
Sacc.cerev  .....CG.-.C-
Acnt.castl  -----

*****
Vann.angli  GGGGGGATTTATATTGGAIGG-TT-AGCGGTGAAATGC-ATGGATCCA-TTCA-AGATAAAC-TAAAGC 1064
Paul.chrom  ....TGC.AG....CAGC.-CC-.A.....T..T...T.G-C.G.-..CT...CT..
Eugl.rotun  ....TGC.AG....CC.C.-CC-.A.....T..T...T.G.-GG.-..CT...CT..
Sacc.cerev  ....C.CGG....CA.T.-C-.A.....T..T...TT.-.G.-..CT...CT..
Acnt.castl  ....C..A....TA.T.-C-.A.....T..T...TT.-.GA.-..T...TCT..

```

**Figure 4.2 continued.**

[illegible]

**Figure 4.2 continued.**

[illegible]

Figure 4.2 continued.

```

Vann.angli
Paul.chrom
Eugl.rotun
Sacc.cerev
Acnt.castl
-----TCGCGGCG

Vann.angli
Paul.chrom
Eugl.rotun
Sacc.cerev
Acnt.castl
-----TATTGTATAGATTATT-TTTAA-AC TTCTT-----AGAGGGACT-- 1545
-----TCTGGAT.GGG.CCG-C-----
-----TGAGATGGC.GG.T-----
-----TTGC.GG--A.CC-----
ACGTCA-TCCCGCCGGCAGGCCCGGTCGGTGGGCG...GG.T.CGGCG-.CCGT-G.....
-----

Vann.angli
Paul.chrom
Eugl.rotun
Sacc.cerev
Acnt.castl
-----AT-TG-GTTT-----C-TAAACCA-----AAGGAAG 1568
-----C-.A.G-----G-.GT-.G-----GT-----
-----C-.GG-----A-.G-.G-----T-----
-----C-.-----A-.G-.G-----T-----
-----GC-.C.CGC-----G-----GC-----

Vann.angli
Paul.chrom
Eugl.rotun
Sacc.cerev
Acnt.castl
-----T-AAAGCGGATAACAGGTC TG-IGATGCCC-TTAG-ATGTTCTGGG-CTGCACGGCGGCTACAATGA-TAATCT-CAACG 1641
.CCG...A.....C.....C...-C.CGT...T.A
.TCG...A.....C.....C...-C.CGT...T.
.TTG...A.....A.C.....C...-CGGAGC-.G.
.TTG...A.....C.....C...-T.ATC-...

Vann.angli
Paul.chrom
Eugl.rotun
Sacc.cerev
Acnt.castl
-----AGTATTATTATT-----
...T.AC.A-----
...CTC.A-----
...C.AA-----
...CCGCT.CAATCGAGGGCGGATGCC-----GTGGG-----GTCAAACCCCAACTG
-----

```

Figure 4.2 continued.

```

****
Vann.angli  -----ATTATATGCTTTATT--GG-AA-----AATA--TGGC-TAATCTTA--AACGATTAT-CGTG--CTT 1705
Paul.chrom  -----C..GGC.C-A-.....GGCC-----G.....TTA..AACG.G-.....-A.G
Eugl.rotun  -----C..G.C-A-.....G.C-.....G.....CTC.....CG.....-A.G
Sacc.cerev  -----C..GGCC--A-G-.....GG.C-..T.G-.....GTG..ACTCCG-.....-G
Acnt.castl  TGTCGCTGTCTCGATCGGC..GGGC--A-T-.....GG.C-C.G-.....TGC..ATT.A-.....-G

*****
Vann.angli  -GGGATAGATGATTGCAATT-ATTTATCTTG-AACGAGGAATTCCTAGTAATC-GTA-TATCATCAGTGTGCGATGATTA 1780
Paul.chrom  -.....T...T...T...G...C-.....A...CG-GG.....ACCC.NAT.....
Eugl.rotun  -.....T...T...T...G...C-.....A...CG-CG.....T.AC.C..AT.....
Sacc.cerev  -.....GC...T...T...GC...C-.....G...C-AG.....CT...T.....
Acnt.castl  -.....C...T...T...G...C-.....G...CG-AG.....CTC...T.....

*****
Vann.angli  CGTCCCTGCCCTTGTACACACCGCCCGTCTGC-TACCGA--TT-GGATGTC-CGG-TGAAATC--CTCGGATAACA 1852
Paul.chrom  .....T.....A.....A.....C.T.T-T.....GGCT--AG...ATGT.
Eugl.rotun  .....T.....A.....A.....T-TA.....GG.T--A...CTGTT
Sacc.cerev  .....T.....AG.....A...CT-TA.....GGC--TCA....CTG-
Acnt.castl  .....T.....C.....A.....A.....T.....--GC

*****
Vann.angli  ATAAATACTTTACTTTGT-----AAAGTTTAT-TGT 1882
Paul.chrom  GCCTGGGG.GGTTGCG-----CACTGCGGGCA-.C
Eugl.rotun  GCCTC.GG..GGT..C-----CTTCCTT.GGCA-.AC
Sacc.cerev  C.T.GAGAAGGGG-GCAAC-----TCCAT--C.CAGA-GCG
Acnt.castl  CGTGGCC--CTACGCAATCCGGCAACCGG-----CTTGTGAGGTC-TCCCTT-.TG-GCG

*****
Vann.angli  TAGAAG--TTGA-TTAA-CCTTA-TCATCTAG-AGGAAGA-AAAA-G-TCGTAACAAGGT-TTCCGTAGGTGAACCT-GCAGATGGATCA---- 1962
Paul.chrom  AT...C---GA-AC.....G-G.G.....T.....A.....A.....
Eugl.rotun  .G.....GAC-CC...T.G.....T.....A.....
Sacc.cerev  G...T-----G-AC...T.GG.....CT.....G.A.....TTA-
Acnt.castl  GC.....C...G.....-C...T.....G-G.....-C.....G.G.....TTA-

```

## Preliminary phylogenetic analyses

Initially, two fast phylogenetic analyses were performed to obtain a preliminary placement for *V. anglica* in the eukaryotic tree and to identify taxa which would be useful as suitable outgroups in succeeding analyses. In the first case, the unaligned *V. anglica* ssrRNA sequence was used in conjunction with the SUGGEST TREE program (Maidak *et al.*, 1996). This program aligns the *V. anglica* ssrRNA with the most similar aligned sequence on the ribosomal database and a FASTDNAML maximum likelihood analysis was carried out with 14 pre-aligned eukaryotic ssrRNA gene sequences from taxa which span the eukaryotic tree. The number of nucleotide positions used in SUGGEST TREE and the weighting each position is allocated was predetermined. The program does not carry out a complete maximum likelihood analysis of all the selected sequences, but it simply places the input sequence into the phylogenetic tree without changing the existing tree. In the second instance, a parsimony analysis was performed with the aligned *V. anglica* ssrRNA sequence and the same set of 14 pre-aligned eukaryotic sequences for the 1488 selected nucleotide positions. The analysis was performed using PAUP software with the tree-bisection-reconnection branch-swapping algorithm to rearrange trees.

The 14 taxa were represented by the microsporidian (*E. hellem*), the trichomonad (*T. foetus*), the schizopyrenids (*V. lobospinosa* and *N. gruberi*), the parasitic amoeba (*E. histolytica*), the anaerobic amoeba (*P. balamuthi*), the plasmodial slime mold (*P. polycephalum*), the cellular slime mold (*D. discoideum*), the acetospora (*H. nelsoni*), a yeast (*S. cerevisiae*), a mammal (*M. musculus*), a diatom (*S. costatum*), a naked amoeba (*A. castellanii*) and an apicomplexa (*S. muris*).

## Estimating the phylogenetic position of *V. anglica* in the eukaryotic tree

To obtain the best estimate of the phylogenetic position of the naked, flat fan-shaped amoeba *V. anglica*, phylogenetic trees were determined using parsimony and maximum likelihood analyses. Both analyses were performed on 37 taxa including *V. anglica* and three probable outgroup taxa, using the same set of 1488 nucleotide positions.

Parsimony analyses were performed using PAUP software (Swofford, 1993). A heuristic search was performed with maximum parsimony optimality criterion to identify the most parsimonious tree i.e the tree which requires the least number of

sequence changes to obtain the observed nucleotide sequence data set. To account for the variation in character state transformations, transversion substitutions were weighted double transition substitutions. Trees were constructed using stepwise addition using several starting taxa. Trees were rearranged using a combination of tree-bisection-reconnection, nearest neighbour interchanges and subtree pruning-regrafting branch swapping methods (Swofford, 1993). The sequence data was bootstrap resampled one hundred times to provide an indication of the confidence of each branch for the most parsimonious tree.

Maximum likelihood analyses were performed using the FASTDNAML program (Olsen *et al.*, 1994), which is based on the DNAML maximum likelihood program from the phylogeny inference programs PHYLIP version 3.5c (Felsenstein, 1981, 1993). The analyses were performed under an evolutionary model which did not assume equal evolutionary rates or equal base frequencies. Options were incorporated to use empirical base frequencies, to weight transversions twice as much as transitions, to randomize the input order of the sequences and for global rearrangements to be made during the tree searching process.

The three outgroup taxa chosen from the preliminary analysis were the parasitic amoeba (*E. histolytica*), the cellular slime mold (*D. discoideum*), and the acetospora (*H. nelsoni*). The remaining taxa including algae, plants, fungi, metazoa, alveolates and stramenopiles were all chosen from crown of the eukaryotic tree. All 37 taxa used in these analyses are marked with an asterisk in Table 4.2.

All trees were viewed and printed using TREEVIEW, a program available for several computer systems which allows the examination of PAUP or PHYLIP trees. The TREEVIEW program is available from Dr Rod Page at the WWW site <http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>.

To examine the potential relationships between *V. anglica* and other taxa, trees were imported into the tree analysis program MACCLADE. Constraint trees were constructed to force relationships between *V. anglica* and other taxa, and a heuristic search was conducted to find the shortest tree. The effect on the overall tree length and the negative natural logarithm (-ln) likelihood was examined.

## Results

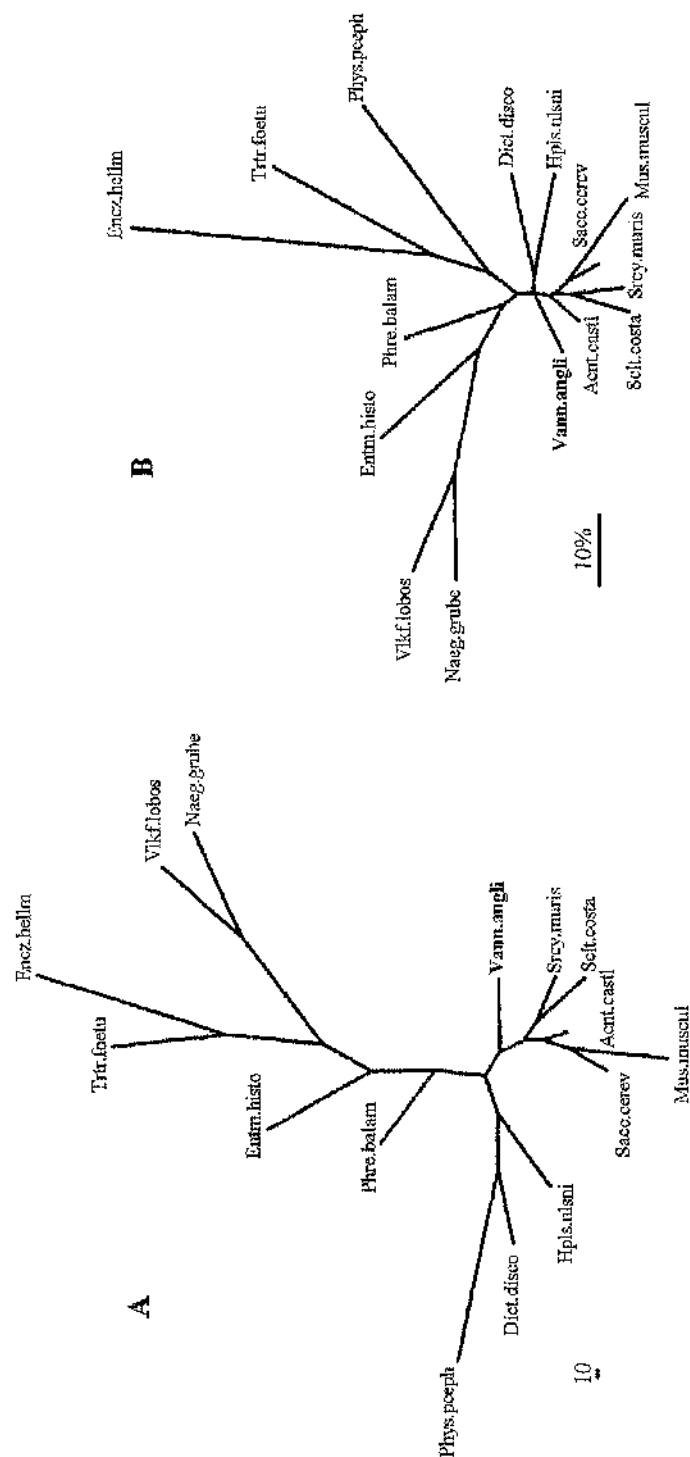
### **Preliminary phylogenetic analyses**

To provide an indication of the phylogenetic position of *Vannella anglica* in the eukaryotic tree, preliminary parsimony and SUGGEST TREE analyses were performed. A range of eukaryotic organisms were used from the most ancient eukaryotic groups, such as the Microsporidia and the trichomonads, through the schizopyrenid amoebae and the slime molds to the more recently diverging groups like the fungi and animals.

The most parsimonious tree and the SUGGEST TREE phylogenies for the range of eukaryotic taxa are shown in Figure 4.3 A and B, respectively. Although, at first glance the trees appear to be quite different, the position from which *V. anglica* branches from the trees is similar. In both instances, *V. anglica* branches from the tree after the microsporidian (*Encephalitozoon hellem*), the trichomonad (*Tritrichomonas foetus*), the schizopyrenid amoebae (*Vahlkampfia lobospinosa* and *Naegleria gruberi*), the anaerobic amoeboid lineages (*Entamoeba histolytica* and *Phreatamoeba balamuthi*), and approximately at the same time or after the divergence of the cellular slime mold (*Dictyostelium discoideum*) and the acetospora (*Haplosporidium nelsoni*). The apicomplexa (*Sarcocystis muris*), the diatom (*Skeletonema costatum*), the naked, lobose amoeba (*Acanthamoeba castellanii*), the fungi (*Saccharomyces cerevisiae*) and the mammal (*Mus musculus*) always branch after *V. anglica*.

Although only the most parsimonious tree is given in Figure 4.3 A, only a few more changes were required to produce the next best nine trees, so they were also considered. Examination of these trees demonstrated that the position of *V. anglica* was stable relative to the other taxa in all cases. The phylogenetic position for many of the other taxa did vary though. There were minor changes in the branching order of the taxa distal to *V. anglica*, where *A. castellanii* would move with respect to *S. muris* and *S. costatum*, to produce the branching order present in the SUGGEST TREE phylogeny. For the more ancient taxa, *E. hellem* and *T. foetus* formed a stable clade, the schizopyrenids formed a stable clade, and the branching order of *E. histolytica* and *P. balamuthi* relative to the schizopyrenids was stable. However, there were significant differences among some taxa for the different trees. The





**Figure 4.3** Preliminary placement of *V. anglica* in the eukaryotic tree. (A) Unrooted most parsimonious radial tree using PAUP. The scale bar indicates 10 observed nucleotide changes. (B) Unrooted maximum likelihood radial tree using SUGGEST TREE. The scale bar indicates ten expected changes every 100 nucleotide positions.

*E. hellem-T. foetus* clade sometimes formed a clade with the plasmodial slime mold *Physarum polycephalum*, a tree topology also seen in the maximum likelihood tree, and *H. nelsoni* branched either near *D. discoideum* or near *P. balamuthi*.

The variations in the tree topologies of these more ancient eukaryotes, however, did not detract from the finding that these preliminary trees clearly demonstrated that *V. anglica* was not related to the schizopyrenid amoebae, the parasitic amoeba *E. histolytica* or the anaerobic amoeba *P. balamuthi*. *V. anglica* appears to branch approximately at or after the divergence of *D. discoideum* and *H. nelsoni*, but before the divergence of the crown taxa which include the multicellular organisms and the contemporary protists.

To provide a more accurate estimation of the position of *V. anglica* further parsimony and maximum likelihood analyses were carried out with the exclusion of highly divergent ancient eukaryotic taxa. Analyses were carried out on 37 taxa including *E. histolytica* which was used as a outgroup, *D. discoideum* and *H. nelsoni* which appeared to branch before *V. anglica*, and 31 taxa covering the alveolates, stramenopiles, naked amoebae, testate amoebae, green algae, cryptomonads, fungi, metazoa and plants.

### **Parsimony analysis**

The most parsimonious phylogram for the 37 eukaryotic taxa with bootstrap resampling values is shown in Figure 4.4. The tree required 3838 evolutionary changes and the -ln likelihood for the tree was 20468.13.

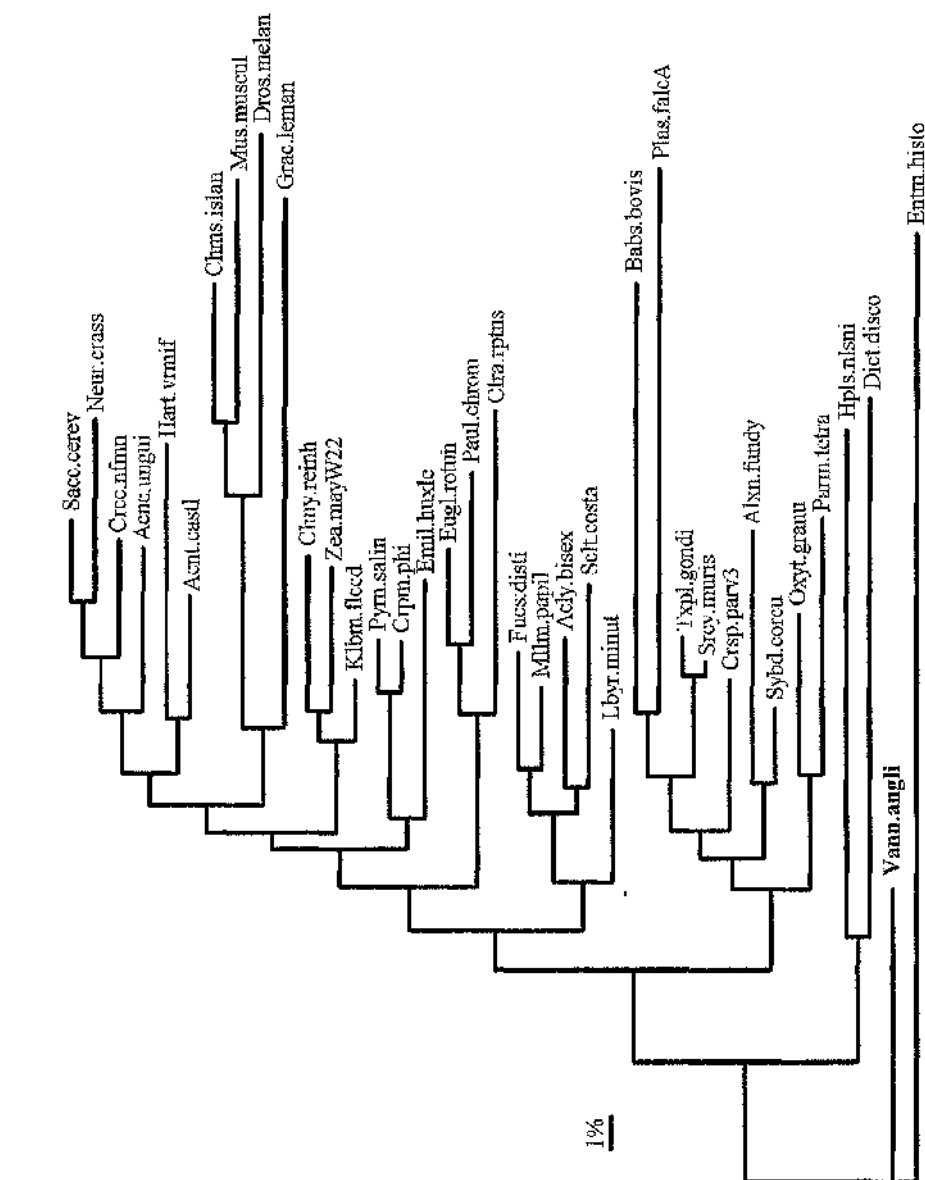
In support of the preliminary analyses, the full parsimony analysis demonstrated that *V. anglica* was the closest outlying lineage to the crown taxa, which included the plants and green algae, metazoa, stramenopiles, fungi, filose testate amoebae, naked lobose amoebae and the alveolates. The bootstrap resampling provides weak support (54%) for the crown taxa with the exclusion of *V. anglica*, and there is 52% support for *D. discoideum* and the defined outgroup *E. histolytica* outlying all the other taxa. There is no bootstrap support for a relationship between *H. nelsoni* and *V. anglica*. All the nine next most parsimonious trees positioned *E. histolytica*, *D. discoideum*, *H. nelsoni* and *V. anglica* in exactly the same position relative to the crown taxa.

Examination of the crown taxa clade demonstrated good bootstrap support for essentially all the established groups. The plants and green algae (*Zea mays*, *Chlamydomonas reinhardtii* & *Klebsormidium flaccidum*) had a bootstrap value of 98%, the cryptomonads (*Cryptomonas phi* & *Pyrenomonas salina*) 100%, the testate amoebae (*Euglypha rotunda* & *Paulinella chromatophora*) 84%, the metazoa (*Mus musculus*, *Chlamys islandica* & *Drosophila melanogaster*) 99%, the naked amoebae (*A. castellanii* & *Hartmannella vermiformis*) 68%, the fungi (*Neurospora crassa*, *Saccharomyces cerevisiae* & *Cryptococcus neoformans*) 98%, the dinoflagellates (*Symbiodinium corculorum* & *Alexandrium fundyense*) 97%, the ciliates (*Paramecium tetraurelia* & *Oxytricha granulifera*) 98%, and the stramenopiles (*Mallomonas papillosa*, *Fucus distichus*, *Skeletonema costatum* & *Achlya bisexualis*) 95%. There was also good bootstrap support for the alveolates (Apicomplexa, dinoflagellates and ciliates) with 77%. Moreover, there was also support for the relationships between the testate amoeba and *Chlorarachnion reptans* with 99%, the fungi and the choanoflagellate (*Acanthocoopsis unguiculata*) with 57%, and between the stramenopiles and the slime net *Labyrinthuloides minuta* with 99% support. There was no bootstrap support for the relationship between the haptophyte *Emiliania huxleyi* and testate amoebae clade or between the red alga *Gracilaria lemaneiformis* and the metazoa, but both these relationships were congruent in all of the nine next best parsimonious trees. There was no bootstrap support for the branching order of the established clades within the crown taxa whatsoever, and the branch lengths separating these groups were small, corresponding from between two and 15 nucleotide changes. Examination of the nine next-best parsimonious trees, however, identified the alveolates and the stramenopiles (including *L. minuta*) as the first two crown taxa to branch from the tree. Differences within these trees were attributed to the variation in the branching order of the remaining clades, and the variable position of *Cryptosporidium parvum* within the alveolate clade.

### Maximum likelihood analysis using FASTDNAML

The maximum likelihood phylogram for the 37 eukaryotic taxa is given in Figure 4.5. The -ln likelihood for the most likely tree was 20444.9.

It was immediately clear that the position of *V. anglica* was different from the position in the most parsimonious tree given in Figure 4.4. In this case, *V. anglica* did not branch just before the crown taxa, it associated with the outgroup taxon



**Figure 4.5** Maximum likelihood phylogram for eukaryotic taxa using the ssrRNA sequence data. The tree was rooted using *Entamoeba histolytica* as an outgroup. The horizontal scale represents one expected evolutionary change every 100 nucleotide positions.

*E. histolytica*, *Haplosporidium nelsoni* and *D. discoideum* form a clade and branch next, followed by the crown taxa. Within the crown taxa all groupings, supported by bootstrap resampling in the most parsimonious tree, were present in the maximum likelihood analysis. Unsupported relationships between the choanoflagellate and the fungi, and the red algae and the metazoa were also present, although the association of the haptophyte and the testate amoebae clade did not. Moreover, in agreement with the parsimony analysis, the alveolates and the stramenopiles were the first two groups among the crown taxa to branch from the tree. The branching order of the remaining crown taxa groups are different from the parsimony phylogeny. However, only small branch lengths, which correspond to no more than one expected change per hundred positions, separate each clade which suggests that the branching order of these clades is uncertain.

### Effect of addition and removal of taxa

To examine the discrepancies between the parsimony and the maximum likelihood phylogenies, parsimony analyses were used to examine the effect that different taxa had on the phylogeny. When the outlying taxon *P. balamuthi* was included in the analysis, the position from which *V. anglica* branched from the tree altered (Figure 4.6 A). *Vannella anglica* branched immediately after *P. balamuthi* and *E. histolytica*, before *D. discoideum* and *H. nelsoni* and the crown taxa. There was weak bootstrap support for *D. discoideum* and *H. nelsoni* forming a clade (55%), and the stronger support for the crown taxa (75%), the branching order of which remained unchanged. The addition of further outgroup taxa provided no more information to the probable position of *V. anglica* in the eukaryotic tree. Indeed, the inclusion of further taxa often yielded misleading results.

The removal of *V. anglica* from the parsimony analysis resulted in *D. discoideum* and *H. nelsoni* forming a clade outlying the crown taxa clade. However, the bootstrap support for these two clades remained poor (Figure 4.6 B). The removal of *H. nelsoni* did not affect the tree topology, but the bootstrap values for the branches within the most parsimonious tree were significant. *Dictyostelium discoideum* associated with the outgroup *E. histolytica*, and *V. anglica* branched next with bootstrap support of 61%. The crown taxa branched next with bootstrap support increasing from 54% with 37 taxa to 90% with the exclusion of *H. nelsoni* (Figure 4.6 C). If *E. histolytica* was removed, the topology of the tree was unchanged and the bootstrap value for the crown taxa increased to 72% (Figure 4.6 D).



Bootstrapped parsimony analysis was also carried out upon subtrees using 13 crown taxa plus *D. discoideum* and *V. anglica*. In turn, one taxa was removed and the effect on the tree topology and the bootstrap values were analysed. The removal of *V. anglica*, *G. lemaneiformis*, *E. huxleyi*, *B. bovis*, *C. parvum*, *C. phi*, *C. neoformans*, *A. unguiculata* and *P. salina* had only minor effects on the tree topology and bootstrap values. The removal of the apicomplexa *P. falciparum* had substantially more profound effects on the subtree. The bootstrap support for *D. discoideum* and *V. anglica* forming an outgroup to the subtree increased from 55% to 97%, the bootstrap value for the alveolate clade increased to 97% and the branching order among the alveolates was clearly resolved with bootstrap values all in excess of 78%.

### Tree rearrangements

As the branch lengths between a majority of the clades were small, only a few evolutionary steps were required to alter the topology of the most parsimonious tree. For instance, when the stramenopiles were made a sister group of the alveolates, the tree length only increased by two, and forcing the naked, lobose amoebae clade (*A. castellanii* & *H. vermiformis*) to be the first to branch after *V. anglica* and the outgroups, only increased the branch length by eight evolutionary steps. However, if the phylogenetic tree was constrained so *V. anglica* formed a clade with the other naked amoebae, the tree length increased by 29 steps. Similarly, in excess of 50 evolutionary changes were required for *V. anglica* to form a clade with the filose testate amoebae. Whereas only 22, 31 and 19 evolutionary changes were required for *V. anglica* to associate with *E. histolytica*, *H. nelsoni* and *D. discoideum* respectively.

## Discussion

### **Preliminary phylogenetic analyses**

Incongruence between the most parsimonious tree and the SUGGEST TREE phylogeny could be attributed to the use of different weighting systems or to the total number of nucleotide positions incorporated in the analyses. However, it is more probable that the limitations of parsimony analysis were being reached. Unlike SUGGEST TREE, parsimony does not account for multiple base substitutions at a single site and thereby assumes a low rate of mutation. Multiple base substitutions are reasonably likely at some of the moderately conserved sites in some of the most ancient eukaryotic taxa, and therefore, parsimony may not perform as well as the SUGGEST TREE method. However, examination of the next best parsimonious trees identified most of the topological features of the SUGGEST TREE phylogeny. Furthermore, the position of *Vannella anglica* remained stable with respect to the neighbouring taxa.

Despite incongruence, the preliminary phylogenetic analyses demonstrated that the naked, lobose amoeba *V. anglica* did not associate with the schizopyrenids *Vahlkampfia lobospinosa* and *Naegleria gruberi*, the plasmodial slime mold *Physarum polycephalum* or the anaerobic amoeboid lineages represented by *Entamoeba histolytica* and *Phreatamoeba balamuthi*.

The separation of *V. anglica* and these other amoeboid lineages is supported by ultrastructural and morphological diagnostic features. *Vannella anglica* has Golgi dictyosomes, branched tubular mitochondrial cristae and non-eruptive movement, but no fruiting bodies or flagellate stages (Page, 1980a). These features clearly distinguish *V. anglica* from the schizopyrenid amoebae which have flattened or discoid mitochondrial cristae, eruptive movement, flagellate forms in most cases, but no dictyosomes (Page & Blanton, 1985). The phylogenetic position of the schizopyrenids suggest that they evolved from an ancient flagellate group, and now the flagella are only present during certain stages or have been lost altogether (Clark & Cross, 1988; Hinkle & Sogin, 1993). The presence of mitochondria also clearly differentiates *V. anglica* from the *Entamoeba* and *Phreatamoeba* lineages which are anaerobic. Tubular mitochondrial cristae are common to both *V. anglica* and the plasmodial slime mold *P. polycephalum*, but fruiting bodies and flagellate



stages are only a trait of the slime mold.

*Vannella anglica* branched from the eukaryotic tree at approximately the same point or just after the cellular slime mold *Dictyostelium discoideum* and the acetospora *Haplosporidium nelsoni*, but before the divergence of the naked, lobose amoebae *Acanthamoeba castellanii* and the other crown taxa. The common possession of Golgi dictyosomes, branched tubular mitochondrial cristae, non-eruptive movement suggests that the preliminary analysis provided a reasonable estimate of the phylogenetic position of *V. anglica*. To improve the accuracy and confidence in the phylogenetic placement of *V. anglica* more thorough analyses were undertaken with fewer outgroup taxa, and more closely related sequences introduced.

### Phylogenetic relationships among the crown taxa

Generally both the complete parsimony and maximum likelihood analyses were congruent with previously published ssrRNA phylogenies for the eukaryotes (Sogin, 1989; Wainright *et al.*, 1993; Schlegel, 1994). Among the crown taxa, maximum likelihood analysis and bootstrap resampled parsimony analysis supported the monophyly of the chlorobionts (plants and green algae), cryptomonads, gymnamoebae, filose testate amoebae and *Chlorachnion*, metazoa, fungi, alveolates (ciliates, dinoflagellates and apicomplexa) and the stramenopiles. These monophyletic groupings are congruent with previous ssrRNA phylogenies (Gunderson *et al.*, 1987; Hendriks *et al.*, 1990; Eschbach *et al.*, 1991; Van de Peer *et al.*, 1992, 1993; Kobayashi *et al.*, 1993; Wilcox *et al.*, 1993; Bhattacharya *et al.*, 1995).

Within the alveolates clade, these results confirm that the ciliates are monophyletic (Bernhard *et al.*, 1995), and a sister-group to the dinoflagellates and the apicomplexans (Leipe *et al.*, 1994). In the ten most parsimonious trees, the position which *Cryptosporidium parvum* occupied varied among the dinoflagellates and the apicomplexa. This result complements previous phylogenies, using partial sequence analysis, which noted that *Cryptosporidia* did not show an especially close relationship with any other apicomplexa (Johnson, 1990).

The strong bootstrap support for the stramenopile assemblage, and the divergence of *Labyrinthuloides minuta* before the oomycete *Achlya bisexualis* and the other stramenopiles supports the hypothesis that the autotrophic stramenopiles evolved

from a heterotrophic flagellate (Leipe *et al.*, 1994). Other photosynthetic taxa with chloroplasts containing chlorophyll a and c, such as the haptophytes, dinoflagellates and cryptomonads, represent separate lineages which must have acquired their chloroplasts independently.

The branching order of the established clades within the crown taxa could not be resolved. For the maximum likelihood analysis the nodes were separated by small branch lengths, and in the parsimony analysis there was no bootstrap support. In general, other ssrRNA phylogenies have been unable to resolve the branching order of the crown taxa (Sogin *et al.*, 1989; Leipe *et al.*, 1994; Bhattacharya *et al.*, 1995), however, a link has been inferred between the metazoa, fungi and choanoflagellates (Kobayashi *et al.*, 1993; Wainright *et al.*, 1993). In this study, the relationship between the choanoflagellate *Acanthoecopsis unguiculata* and the fungi was detected in the maximum likelihood analysis, and the parsimony analysis where there was weak bootstrap support, although there was no relationship detected with the metazoa.

The branching of the alveolates and the stramenopiles as the first two groups of crown taxa to diverge from the maximum likelihood tree and the ten most parsimonious trees is probably not significant, although these two groups have been shown to branch together previously (Gunderson *et al.*, 1987; Wainright *et al.*, 1993). Distance matrix phylogenies have suggested that the first two crown taxa to diverge from the eukaryotic tree are the red algae and the metazoa (Van de Peer *et al.*, 1993; Embley *et al.*, 1994). An association between the red algae and the metazoa was detected in both the maximum likelihood and parsimony analyses, although it was not supported by bootstrap resampling. Other phylogenies have not detected this association, and several other taxa such as the alveolates and the stramenopiles have been shown to branch after the red algae and before the metazoa (Schlegel, 1994). Moreover, it has been suggested that the red algae diverged later among the crown taxa, at about the same time as the stramenopiles or the chlorobionts (Bhattacharya *et al.*, 1990). The association, and the early divergence of the red algae and the metazoa may be artifacts which become apparent when quickly evolving *Gracilaria* and metazoan sequences are used. In combination with distance matrix methods, there may be long branch attraction between these sequences and divergent outlying taxa (Hendy & Penny, 1989).

### Phylogenetic position of *Phreatamoeba balamuthi*

Interestingly, in both preliminary analyses and when *P. balamuthi* was included in a complete parsimony analysis, it branched before *D. discoideum*. This contrasts with the only other published phylogeny using *P. balamuthi*, a distance matrix method which suggested that it diverged after *D. discoideum* before the evolution of the crown taxa (Hinkle *et al.*, 1994). The deeper branching of *P. balamuthi* reported here, however, supports ultrastructural studies which have shown that this amoeba does not have mitochondria or Golgi dictyosomes. Currently *Phreatamoeba* is classified within the ancient eukaryotic kingdom, Archezoa, on the basis of the absence of Golgi dictyosomes and mitochondria (Cavalier-Smith, 1993). However, the inclusion of *Phreatamoeba* in this taxon, along with the diplomonads and microsporidians, infers that this lineage diverged before the schizopyrenida which is contradictory to both ssrRNA phylogenies. Negative characters, such as the absence of Golgi dictyosomes, must be used with caution. The reduction of dictyosomes could have occurred in much the same way as the selective loss of mitochondria in parasitic lineages. Alternatively, characteristic dictyosomes may not be as obvious in *Phreatamoeba* and may have been overlooked. No dictyosomes were originally detected in the ultrastructural examinations of *Entamoeba* (Gicquaud, 1979), although this has been contested (Cavalier-Smith, 1993). The ssrRNA phylogenetic examination of the *Pelomyxa*, which has also been placed within the kingdom Archezoa, will be of great interest as it may indicate the value of dictyosomes, or their absence, as phylogenetic markers.

### Phylogenetic position of *Vannella anglica*

In line with the preliminary results, the complete analyses suggested *V. anglica* branched at around the same point as the cellular slime mold *D. discoideum*. The phylogenetic position of *V. anglica* was not conserved for both the maximum likelihood and the parsimony analyses. In the maximum likelihood analysis, *V. anglica* branched before the cellular slime mold *D. discoideum* and the acetospora *H. nelsoni*, and the crown taxa which branched next. In the parsimony analysis *V. anglica* branched after *D. discoideum* and *H. nelsoni*, but before the crown taxa. This approximate position of *V. anglica* is much earlier than might have been expected, since *V. anglica* is in the same subclass as *Acanthamoeba castellanii* and *Hartmannella vermiformis* which branched together among the crown taxa.

There are numerous factors which could have influenced the branching position of *V. anglica*. Parsimony phylogenies are particularly susceptible to error if the fundamental assumption of low rate of mutation is violated (Felsenstein, 1978). This can become apparent if highly variable nucleotide positions are included, or if taxa are used which have accumulated substitutions at a high rate or over a long time scale (Felsenstein, 1988).

In the analyses reported here, all hyper-variable nucleotide positions were excluded and a total of 1488 positions were included. As many as 1570 and 1644 nucleotide positions have been used in previous parsimony analyses to examine the phylogenetic position and the relationships between the ciliates (Bernhard *et al.*, 1995), and the algae (Medlin *et al.*, 1994), respectively. The number of sites used in this analysis was therefore reasonably conservative.

To ensure that multiple substitutions were kept to a minimum, the number of the distantly related taxa was limited. The most ancient taxon included in the analysis was *E. histolytica*, and when this taxon was excluded from the analysis the topology of the tree remained unchanged, and support for the crown taxon without *V. anglica* actually improved.

A high mutation rate could be introduced into a sequence by misincorporation of nucleotides by *Taq* DNA polymerase during the polymerase chain reaction. To identify such errors, PCR amplifications can be sequenced directly (Embley, 1991) or they can be cloned and multiple clones can be sequenced (Gunderson *et al.*, 1986). In the case of *V. anglica*, the complete sequence was determined from a single clone, but analysis of multiple clones of partial *ssrRNA* sequences in Chapter 5 estimated the combined PCR and sequence error rate was less than  $2.5 \times 10^{-3}$  for 30 cycles of PCR. This could correspond to as many as four or five errors throughout the whole of the gene. However, since all suspected errors were excluded from the analyses and the variable regions, where errors were most likely to go undetected, were also excluded, it is unlikely that any errors are present in the 1488 nucleotide dataset. Therefore, the possibility that errors could have affected the phylogeny can be discounted.

The phylogeny of a tree may also be influenced by the number of taxa, or individual taxa (Felsenstein, 1988). Parsimony analyses on subtrees indicated that the position

of *V. anglica* was not influenced by the large number of crown taxa. Only two taxa, the apicomplexan *P. falcifarum* and the acetospora *H. nelsoni* appeared to have any significant influence on the outcome of phylogenies. The removal of the *P. falcifarum* from a subtree increased the support for the crown taxa from 55% to 97%. *Plasmodium falcifarum* appeared to be attracted to the outlying *V. anglica* and *D. discoideum* in the subtree. This phenomenon is characteristic of taxa with a long branch length and a relatively high rate of sequence evolution (Felsenstein, 1988). However, when this taxon was removed from the complete parsimony analysis, there was only a marginal increase in the support for the crown taxa. This is presumably because in the complete analysis the inclusion of more alveolate taxa stabilised its position. The long branch length attraction of *Plasmodium* species has been evident in other ssrRNA phylogenies which have shown *Plasmodium* to branch deeper in the tree than other crown taxa (Sogin *et al.*, 1989). The removal of *H. nelsoni* had a considerable effect on the parsimony phylogeny using the complete dataset. The support for the crown taxa clade increased from 54% to 90%, and the support for *V. anglica* branching immediately before the crown taxa was 60%. Previous phylogenies have also shown good support for the crown taxa branching after the divergence of *D. discoideum* (Leipe *et al.*, 1993; Bhattacharya *et al.*, 1995), but in each of these cases an acetospora taxon was not included in the analysis.

Phylogenies may also be affected by unequal base frequencies resulting in misleading phylogenies. The only reported examples of this phenomenon have involved the examination of ancient eukaryotic lineages which have large differences in their base compositions (Sogin *et al.*, 1989; Leipe *et al.*, 1993). However, the *V. anglica* ssrRNA has a low G+C content of 37%, compared to 45-55% for most other taxa in the analysis, and this base composition bias may have influenced the phylogeny. However, the base bias was mostly associated with the hyper-variable regions which were excluded for the analysis. Furthermore, since the evolutionary model in the maximum likelihood analysis did not assume that the base frequencies were equal, this factor should not have affected the phylogeny.

The exact position from which *V. anglica* branches from the eukaryotic tree remains unclear. Since parsimony analysis is most reliable with few outgroups, and the reverse is true for maximum likelihood, the best estimation for the branch point of *V. anglica* is probably after *D. discoideum* and before the crown taxa. To clarify the branching position of *V. anglica*, sequences from more taxa which branch in this

region may be required to overcome the sampling effects of individual sequences such as *H. nelsoni*. Alternatively, maximum likelihood analysis using a combination of a more restrictive dataset and more distantly related taxa may provide an improved indication of the branching order in this region of the tree.

Despite the predominant use of the *ssrRNA* for inferring phylogenies, there is some evidence from protein data that suggests that *rRNA* phylogenies may be inaccurate. Generally phylogenies inferred from protein data are largely congruent with *rRNA* phylogenies. However, with protein datasets *E. histolytica* tends to branch deeper in the eukaryotic tree, whereas *D. discoideum* branches much later along with the crown taxa (Hasegawa *et al.*, 1993). A recent protein phylogeny using 19 different protein datasets showed that *D. discoideum* has a closer affinity to animals and fungi, than to plants (Kuma *et al.*, 1995). This analysis, however, only used one or two species from each group, and the use of ancient outgroups such as bacteria is controversial. Nevertheless, *ssrRNA* must not be regarded as the only basis for inferring phylogenies, but as a useful tool which can be used in conjunction with protein, morphological and ultrastructural data for examining evolutionary relationships among organisms.

### **Polyphyly within the phylum Rhizopoda**

In an effort to reduce the polyphyly within the Rhizopoda, several independent lineages have been identified and removed from the phylum of six classes previously proposed (Page, 1987). In most cases, only two classes, the Lobosea and Filosea, remain in the phylum Rhizopoda (Schuster, 1989; Cavalier-Smith, 1993).

This study complements a previous *ssrRNA* phylogeny which demonstrated that the Rhizopoda is still polyphyletic. The lobose, naked amoebae *A. castellanii* and *Hartmannella vermiformis* (class Lobosea, subclass Gymnamoebia) and the filose testate amoebae, *Paulinella chromatophora* and *Euglypha rotunda* (class Filosea), represent quite separate lineages (Bhattacharya *et al.*, 1995). The phylogeny presented here also demonstrated that the naked, lobose amoeba *V. anglica* (class Lobosea, subclass Gymnamoebia) did not associate with either the gymnamoebae, the filose testate, or any other members of the crown taxa. Manual rearrangements of the most parsimonious tree to force *V. anglica* to branch with either of these amoeboid lineages required relatively large increases in the total branch length. These results thereby suggest that *V. anglica* represents a third amoeboid lineage

within the class Lobosea, and the second within the subclass Gymnamoebia.

Previously the *Entamoeba* was regarded as a member of the Gymnamoebia (Page 1976, 1987), but the absence of mitochondria, the presence of endosomes and multinuclear cysts, and ssrRNA phylogeny have resulted in this taxon being removed from this subclass (Cavalier-Smith, 1993). In contrast, despite the clear separation of *V. anglica* from *A. castellanii* and *H. vermiformis* in the ssrRNA phylogeny, there is no single obvious morphological or ultrastructural feature which can be used to distinguish *Vannella* from the both the other two taxa. *Vannella* is more similar to *Acanthamoeba* with respect to the flattened locomotive forms, the hyaline, radiating pseudopodia present in most floating forms and the general absence of cytoplasmic crystals, although *Acanthamoeba* have centrioles and *Vannella* do not. In contrast, the cell surface glycostyles of *Vannella* and *Hartmannella* appear to be more differentiated than those of *Acanthamoeba*. The only apparent diagnostic features which separate *Vannella* from *Acanthamoeba* and *Hartmannella* are the flattened flabellate, spatulate or ovoid locomotive morphologies with a substantial hyaloplasm, and the absence of cysts. The absence of cysts in *Vannella* is probably not significant because cysts are present in freshwater species of the morphologically similar *Platyamoeba*. Using the gross morphology features as a phylogenetic feature, it can be hypothesised that *Vannella* would form a clade with most other members of the Vannellidae, Thecamoebidae, Flabellulidae and Hyalodiscidae with the exclusion of all other gymnamoebae.

In the absence of a single majestic character which supports the separation of *Vannella* from *Acanthamoeba* and *Hartmannella*, the diversity exhibited among the gymnamoebae ensures that any character chosen to subdivide the subclass Gymnamoebia will be a very subjective choice. However, with the addition of more ssrRNA sequences, and detailed ultrastructural studies, clear relationships among the gymnamoebae may become evident, and useful phylogenetic features may be identified. This should stimulate the development of a classification scheme based upon evolutionary relationships rather than subjective interpretations of comparative morphology.

## **CHAPTER 5**

### **Identification and phylogenetic analysis of morphologically similar fan-shaped amoebae using partial ssrRNA sequence data**



## Introduction

### **Taxonomy of the fan-shaped amoebae**

In 1926, Schaeffer described four fan-shaped or flabellate amoebae which were grouped in the genus *Flabellula*. The four members of this genus, *F. citata*, *F. crassa*, *F. mira* and *F. pellucida* were described as having a triangular or fan-shaped locomotive form with a broad anterior. No pseudopodia are formed during locomotion, although short blunt scars may appear along the anterior edge. Deep rifts may occur in the anterior hyaloplasm, but these are either promptly filled or the hyaloplasm at the side of the rift gradually moves back towards the posterior end. A trailing, thin root-like uroid is also frequently present. The floating forms have radiating pseudopodia which vary in size from about one half to several times the length of the central mass. *Flabellula citata* was designated the type species. (Schaeffer, 1926).

In 1965, Bovee reassessed the characteristics of *Flabellula* and proposed that some members of this genus, including the type species *F. citata*, do have pseudopodia during locomotion. The description of the genus was amended to include only fan-shaped amoebae often with uroidal filaments and sometimes with round-tipped pseudopods during locomotion, and a floating form with pseudopodia two or more times longer than the body mass (Bovee, 1965). On account of these changes to the genus, early descriptions of two fan-shaped amoebae with pseudopodia, *F. calkinsi* and *F. patuxent* (Hogue, 1914, 1921), were included in this genus. Bovee also proposed a second genus, *Vannella*, to accommodate fan-shaped amoebae without pseudopodia during locomotion and trailing uroidal filaments, although old radiate pseudopodia may trail until retracted (Bovee, 1965). The isolate *F. mira* was made type species of this genus, however following confusion over the habitat of this isolate and its resemblance to a common freshwater isolate *V. miroides* (Bovee, 1965), the name and type species *V. mira* has been replaced by *V. miroides* (Page, 1983), despite apparent morphological differences. Bovee assigned many species to the *Vannella* genus, including *F. crassa*, *Rugipes vivax*, *Thecamoeba orbis*, *Hyalodiscus elegans* and *H. caeruleus* (Schaeffer, 1926), and *H. simplex* (Wohlfarth-Botterman, 1960), most of which were later removed. As a result of collections around the United Kingdom, several marine species have been added to this genus (Page, 1979a, 1980a), but, to accommodate these new marine species, the genus was

amended to include amoebae with short or non-radiating pseudopodia (Page, 1979a).

In 1969, Page erected a new genus *Platyamoeba* for a tongue-shaped or linguiform freshwater isolate which he called *P. stenopodia*, and a fan-shaped, but otherwise similar freshwater isolate, *P. placida*, which was previously described as a member of the genus *Rugipes* (Page, 1968, 1969). Members of the genus *Platyamoeba*, to which *P. placida* was designated the type species, are described as flattened, with a usually ovoid, truncately elliptical or linguiform outline, usually with a length: breadth ratio larger than one. No pseudopodia are present during locomotion, although longitudinal, lateral folds or wrinkles are common. The floating forms are generally rounded with short or long, blunt hyaline pseudopodia. Known cysts are spherical with a two layered wall (Page, 1969). Since 1969 several species, almost all from the marine habitat, have been assigned to the genus *Platyamoeba* (Page, 1974a; Sawyer, 1975b).

The common overall shape of 'fan-shaped' amoebae was not a sufficiently rigid character to allow the classification of *Vannella*, *Flabellula* and *Platyamoeba* together. The *Vannella* and *Flabellula* genera were classified in different subfamilies of the family Flabellulidae, on account that *Vannella* species did not produce anterior hyaloplasmic pseudopodia or trailing uroidal filaments (Bovee, 1970). The *Platyamoeba* genus was placed in a separate family, Thecamoebidae (Page, 1969), due to the presence of wrinkles which were superficially similar to the parallel ridges present in *Thecamoeba* species.

In 1974, the *Platyamoeba* genus was temporarily moved to a newly erected family Striamoebidae with the type genus *Striamoeba* (Jahn *et al.*, 1974), however, the *Striamoeba* genus was rejected as this name applied to some members of the *Thecamoeba* (Page, 1977). In 1976, the *Platyamoeba* genus was returned to the family Thecamoebidae with *Thecamoeba*, the newly erected genera *Pessonella* (Pussard, 1973) and *Sappinia* (Goodfellow *et al.*, 1974), and *Vannella* (Page, 1976). However, to accommodate the *Vannella* and *Pessonella* genera the importance of the ridges in the family description was clearly compromised.

In 1975, a new genus of fan-shaped amoebae, *Clydonella*, was proposed to accommodate the marine isolate, *Rugipes vivax* (Schaeffer, 1926; Sawyer, 1975a). In an accompanying paper several other *Clydonella* species were described from the

surface waters of Chincoteague Bay, Virginia, USA (Sawyer, 1975b). Despite the similarities with *Platyamoeba* and *Vannella*, Sawyer described *Clydonella* as quite distinct from these genera principally because in the floating form the pseudopodia of *Clydonella* fork slightly at the tips (Sawyer, 1975a, 1975b). From Sawyer's observations, Page described *Clydonella* as having "light microscopical characters which appear intermediate between those of *Vannella* and *Platyamoeba*" (Page, 1983). Interestingly, although Page and Sawyer have isolated *Platyamoeba*, Page has only described *Vannella* species and never *Clydonella*, whereas Sawyer primarily describes *Clydonella* and rarely *Vannella*, despite sampling comparable habitats.

Following the advent of transmission electron microscopy (TEM), the most recent classification of the naked amoebae has placed considerable emphasis on the cell surface structure as a taxonomic character (Page, 1987). In this scheme, *Vannella* and *Platyamoeba* have been removed from the Thecamoebidae and placed in a new family, Vannellidae in the order Euamoebidae (Page, 1987). *Flabellula* remains in the family Flabellidae, but is in a separate order, Leptomyxida.

Over the years, conflicting classification schemes which have been proposed for the naked amoebae have served to demonstrate that relationships have remained poorly understood, and that there is little indication of which characters provide the most accurate phylogenetic markers (Schaeffer, 1926; Chatton, 1953; Bovee & Jahn, 1966; Singh & Das, 1970; Jahn *et al.*, 1974; Page, 1976; Levine *et al.*, 1980; Page, 1987). The current classification of the subclass Gymnamoebia is shown in Table 5.1. This subclass includes all the non-sporulating, naked amoebae except the schizopyrenids which are clearly a distinct and unrelated group (Page & Blanton, 1985; Clark & Cross, 1988). Despite the morphological similarities, Page has rejected the *Clydonella* genus in the absence of examination of the cell surface structure (Page, 1987). It has been included in the classification presented here as a genus *incertae sedis*.

**Table 5.1** The classification of the naked amoebae (modified from Page, 1987; Page & Siemensma, 1991; Smirnov & Goodkov, 1993).

**Subclass Gymnamoebia Haeckel, 1862**

Order Acanthopodida

Family Acanthamoebidae

*Acanthamoeba*, *Protoacanthamoeba*.

Order Euamoebida

Family Amoebidae

*Amoeba*, *Chaos*, *Deteramoeba*, *Hydramoeba*,  
*Parachaos*, *Polychaos*, *Trichamoeba*.

Family Hartmannellidae

*Cashia*, *Glaeseria*, *Hartmannella*, *Saccamoeba*.

Family Thecamoebidae

*Dermamoeba*, *Paradermamoeba*,  
*Pseudothecamoebea*, *Sappinia*, *Thecamoeba*,  
*Thecochaos*.

Family Vannellidae

*Pessonella*, *Platyamoeba*, *Vannella*.

Family Paramoebidae

*Dactlyamoeba*, *Mayorella*.

Family Vexilliferidae

*Neoparamoeba*, *Pseudoparamoeba*, *Vexillifera*.

Order Leptomyxida

Suborder Leptoramosina

Family Gephyramoebidae

*Gephyramoeba*.

Family Stereomyxidae

*Stereomyxa*.

Suborder Rhizoflabellina

Family Flabellulidae

*Flabellula*, *Paraflabellula*.

Family Leptomyxidae

*Leptomyxa*, *Rhizamoeba*, *Ripidomyxa*.

Order Loboreticulatida

Family Corallomyxidae

*Corallomyxa*.

**Uncertain families within Gymnamoebia**

Family Echinamoebidae

*Comandonia*, *Echinamoeba*, *Filamoeba*.

Family Entamoebidae

*Entamoeba*.

Family Hyalodiscidae

*Hyalodiscus*, *Flamella*.

**Uncertain genera within Gymnamoebia**

*Bovee*, *Clydonella*, *Unda*.

## Taxonomic characters for identification

Members of the Vannellidae family have 'fan-shaped' locomotive forms without pseudopodia or uroidal elements. Genera belonging to this family include *Vannella*, *Platyamoeba*, *Pessonella*, and arguably *Clydonella* too (Page, 1987, 1988; Sawyer, 1975a). A listing of all the described members of these genera as well as unidentified isolates described in Chapter 2, are presented in Table 5.2. Details of their habitat, taxonomic characters and appropriate references are also given.

Freshwater isolates are reasonably easy to identify. Generally members of the *Vannella* genus have long tapering pseudopodia in the floating form, and no cysts have been detected. Cysts have been detected for *V. cutleri*, although the absence of pseudopodia indicates that this isolate should not have been designated to this genus (Page, 1988). All described freshwater *Platyamoeba* produce cysts, although this ability may be lost after a prolonged period in culture. The pseudopodia of the floating form are occasionally long, but are blunt. Light microscopical observations of the floating and locomotive forms indicate that *Pessonella marginata* closely resembles *Platyamoeba* species, although no cysts have been observed (Page, 1988). The acceptance of this monotypic genus therefore remains uncertain.

The distinction between marine isolates of the *Vannella* and *Platyamoeba* genera, and the related genus *Clydonella*, is far more problematic. Page described some marine isolates of these genera as "virtually indistinguishable with the light microscope" (Page, 1983). The presence of wrinkles, which often appear in the hyaloplasm of *Platyamoeba* isolates in the locomotive form, is regarded as an important diagnostic feature for distinguishing *Vannella* and *Platyamoeba*, although these wrinkles are usually transient, and can occur in some *Vannella* species when they change direction.

The shape of the locomotive form is perhaps the most important morphological character for the identification of marine *Vannella* and *Platyamoeba* isolates. Numerous terms have been used to describe the general shape of these amoebae such as fan-shaped or flabellate, oblong, oval, semi-circular, spatulate, triangular and tongue-shaped or linguiform. Moreover, these terms are qualified by a whole selection of terms such as "usually", "sometimes", "occasionally" and "more often than not". The shape of amoebae may vary considerably within a population of cells; the variable shape of the locomotive form of *Vannella* isolates is even considered

**Table 5.2** Listing of morphologically similar *Vannella*, *Platyamoeba*, *Pessonnella* and *Clydonella* species and unidentified isolates.

Isolate	Habitat	Greatest dimension	L:B ratio	Floating form pseudopodia	Cysts	Nucleolar material	Surface structure	Reference
<i>Platyamoeba australis</i>	M	32 µm	1.0	L	Absent	Central	Glycocalyx	Page, 1983
<i>P. bursella</i>	M	19-25 µm	1.0	S	Absent	Central	Glycocalyx	Page, 1974a
<i>P. douvresi</i>	M	13 µm	1.0	S	Unknown	Central	Unknown	Sawyer, 1975b
<i>P. calycinucleolus</i>	M	32-39 µm	0.8-0.9	L	Absent	Central	Glycocalyx	Page, 1974a
<i>P. langue</i>	M	9 µm	1.1	S	Unknown	Central	Unknown	Sawyer, 1975b
<i>P. lei</i>	M	20 µm	1.3-1.4	A	Unknown	Central	Unknown	Sawyer, 1975b
<i>P. mainensis</i>	M	26 µm	1.1	M	Absent	Central	Glycocalyx	Page, 1971
<i>P. murchelanoi</i>	M	11 µm	1.2	A	Unknown	Central	Unknown	Sawyer, 1975b
<i>P. placida</i>	F/W	25 µm	1.2	S-M	Present	Central	Glycocalyx	Page, 1969
<i>P. plurinucleolus</i>	M	11-26 µm	1.0-1.1	M-L	Absent	Parietal	Glycocalyx	Page, 1974a
<i>P. schaefferi</i>	F/W	27-29 µm	>1	L	Present	Unknown	Unknown	Singh & Hanumiah, 1979
<i>P. stenopodia</i>	F/W	24 µm	2.5	L	Present	Central	Glycocalyx	Page, 1969
<i>P. weinsteini</i>	M	13 µm	0.9	M	Unknown	Central	Unknown	Sawyer, 1975b
<i>Vannella aberdonica</i>	M	9 µm	1.0	A	Absent	Central	Glycostyles	Page, 1980a
<i>V. arabica</i>	M	27 µm	0.8	S	Absent	Central	Glycostyles	Page, 1980a
<i>V. anglica</i>	M	21-24 µm	1.0-1.2	S-M	Absent	Central	Glycostyles	Page, 1980a
<i>V. caldonica</i>	M	16 µm	1.2	A	Absent	Central	Glycostyles	Page, 1979a
<i>V. cirrifera</i>	F/W	25-30 µm	1.1	L	Absent	Unknown	Unknown	Page, 1988
<i>V. crassa</i>	M	50-75 µm	1.0	S-M	Unknown	Unknown	Unknown	Bovee, 1965
<i>V. cutleri</i>	F/W	27 µm	1.2	A	Present	Unknown	Unknown	Singh & Hanumiah, 1979
<i>V. devonica</i>	M	22 µm	0.9	L	Absent	Parietal	Glycostyles	Page, 1979a
<i>V. lata</i>	F/W	33 µm	0.6	L	Absent	Central	Glycostyles	Page, 1988
<i>V. miroides</i>	F/W	25-35 µm	0.8-1.0	L	Unknown	Central	Unknown	Bovee, 1965
<i>V. platypodia</i>	F/W	16-21 µm	0.9	L	Absent	Central	Glycostyles	Page, 1968
<i>V. sensilis</i>	M	15-24 µm	0.9	S	Unknown	Unknown	Glycostyles	Bovee & Sawyer, 1979
<i>V. septentrionalis</i>	M	21-23 µm	0.8-0.9	L	Absent	Central	Glycostyles	Page, 1980a
<i>V. simplex</i>	F/W	42-52 µm	0.8	L	Absent	Central	Glycostyles	Bovee, 1965

Table 5.2 Continued.

Isolate	Habitat	Greatest dimension	L:B ratio	Floating form pseudopodia	Cysts	Nucleolar material	Surface structure	Reference
<i>Clydonella rosenfield</i>	M	17 $\mu$ m	1.0	M-L	Unknown	Central	Unknown	Sawyer, 1975a
<i>C. sindermanni</i>	M	28 $\mu$ m	0.9	L	Unknown	Central	Unknown	Sawyer, 1975a
<i>C. vivax</i>	M	14 $\mu$ m	1.2	L	Unknown	Central	Unknown	Sawyer, 1975b
<i>C. wardi</i>	M	18 $\mu$ m	1.1	L	Unknown	Central	Unknown	Sawyer, 1975a
<i>Pessonnella marginata</i>	F/W	45 $\mu$ m	0.8	M	Absent	Parietal	Unknown	Pussard, 1973
SIA	F/W	25 $\mu$ m	1.1	S-M	Present	Unknown	Unknown	Chapter 2
Species E	M	10 $\mu$ m	1.5	S-M	Absent	Unknown	Unknown	Chapter 2
VP3	M	22 $\mu$ m	0.8	M-L	Absent	Unknown	Unknown	Chapter 2

With respect to habitat, M = marine and F/W = freshwater, and for the floating form pseudopodia, A, S, M and L represent absent, short, medium and long pseudopodia, respectively. The values for the greatest dimensions and length: breadth ratios (L:B) are averages.

Several of the species presented here have been described under different names. *Platyamoeba placida* was formerly known as *Rugipes placida* (Page, 1968), and *P. leei* was formerly known as *Lingulamoeba leei* (Sawyer, 1975a). *Vannella cirrifera* was formerly known as *Amoeba mira*, *Flabellula mira* (Schaeffer, 1926) and *V. mira* (Bovee, 1965). *Vannella crassa* was formerly known as *F. crassa* (Schaeffer, 1926). *V. platypodia* was known as *F. platypodia* (Page, 1968), and *V. simplex* was known as *Hyalodiscus simplex* (Wohlforth-Botterman, 1960). Finally *Clydonella vivax* was formerly known as *Rugipes vivax* (Frenzel, 1892) and *V. vivax* (Bovee, 1965).

to be a diagnostic character (Page, 1983).

To help quantify the shape of amoebae, the length: breadth (L:B) ratio can be used as an index. The range of the L:B ratio provides an indication of the extent of variation within a population of cells, but often only an average L:B ratio is given in species descriptions. The use of an average L:B ratio is useful for delimitating values for phylogenetic purposes, but a significant amount of information regarding the shape dynamics is lost. In some cases, the average size of the greatest dimension is a useful diagnostic character. The greatest dimension of those species described varies from approximately 10  $\mu\text{m}$  in *V. aberdonica*, *P. langae* and *P. murchelanoi* to approximately 50  $\mu\text{m}$  or more for *V. simplex* and *V. crassa*. However, the majority of described isolates have their greatest dimension between 20 and 30  $\mu\text{m}$ , and considerable size variation has been identified among strains of *P. plurinucleolus* (Page, 1974a). Therefore, it is clear that shape and size are not rigid diagnostic characters, especially since both may vary considerably with the physiological state of the cell population, culture conditions and the rate of movement. Certain guidelines have been suggested to try to standardize measurements and observations of amoebae (Page, 1983), but these problems remain largely unresolved.

The size and shape of pseudopodia of floating amoebae are useful for distinguishing between *Vannella* and *Platyamoeba*. The pseudopodia of freshwater *Vannella* are generally long and taper to a tip, whereas the pseudopodia of freshwater *Platyamoeba* do not taper. However, the pseudopodia of marine *Vannella* do not taper either, and there is considerable variation in the length within both genera. The length of the pseudopodia is usually described with respect to the size of the central mass. Some *Vannella* and *Platyamoeba* isolates, such as *V. aberdonica* and *P. leei* do not have radiating pseudopodia, others, including *V. anglica* and *P. placida*, have pseudopodia up to the length of the central mass, whereas some freshwater isolates have pseudopodia two or three times longer than the diameter of the central mass.

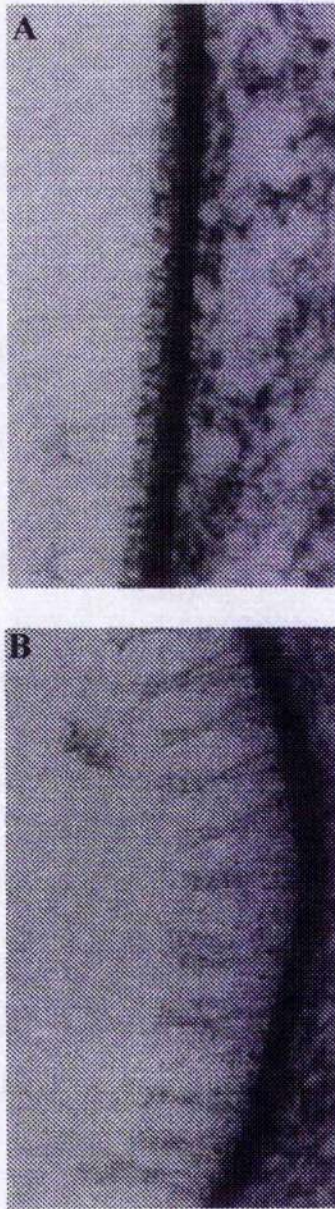
In the absence of rigid morphological characteristics, it may be difficult to confidently assign a flat, fan-shaped amoeba to the *Vannella* or *Platyamoeba* genera, particularly if the isolate is marine. However, the difficulty is compounded if another morphologically similar genus, *Clydonella*, is considered. There are only a



few records on the light microscopical observations of *Clydonella* (Sawyer, 1975ab), and members of this genus appear to be essentially indistinguishable from marine *Vannella* and *Platyamoeba* (Page, 1983).

Examination of the arrangement of the nuclear material has been used as an aid to the identification of some isolates. Epifluorescence light microscopy with the DNA specific fluorochrome DAPI or transmission electron microscopy of the nucleus has shown that the majority of amoebae have a nucleus with a prominent central nucleolus, but in the cases of *P. plurinucleolus*, *V. devonica* and *Pessonnella marginata*, the nucleolar material is arranged into several parietal lobes. Mitotic behaviour is also considered to be a useful character for identification. During mitosis there are no centrioles or other microtubular organising centres and the nucleolus disappears during prophase, but the timing and extent to which the nuclear membrane disintegrates are variable. In *P. stenopodia* and *P. placida*, disintegration takes place at an early stage and this property was used to distinguish *Platyamoeba* from *Vannella* species (Page, 1968, 1969). Subsequently, it has become clear that the nuclear membrane may also disintegrate quite early during metaphase in *V. aberdonica* and during anaphase in *V. arabica*. Alternatively, remnants of the nuclear membrane persist and form polar caps as in *V. septrentionalis* (Page, 1980a). Similar variation in the mitotic behaviour has been detected in *Platyamoeba* species (Page, 1971, 1974a). Details of the mitotic behaviour for most fan-shaped isolates is not available because the fate of the nuclear membrane during any particular stage in mitosis is difficult to assess.

Since the advent of the electron microscope, characteristics of the cell surface structure have been used to classify fan-shaped amoebae into the *Vannella* or *Platyamoeba* genera. Examination by TEM has shown that the cell surface structure, or glycocalyx, is composed of a dense basal layer of approximately 5-10 nm and a fibrous layer of tightly packed indistinct elements of approximately 10-40 nm. Alternatively, the glycocalyx may have distinct glycostyles extending 90-130 nm from the plasma membrane (Page & Blakey; 1979; Page, 1979a, 1980ab). Isolates with a "fuzzy" glycocalyx are classified as *Platyamoeba* and isolates with distinct glycostyles are classified as *Vannella* (Page 1983). Figures 5.1 A and B show ultrastructural details of the cell surface from characteristic *Platyamoeba* and *Vannella* isolates.



**Figure 5.1** Electron micrographs of cell surface structure from characteristic *Platyamoeba* and *Vannella* species (Reproduced from Page, 1983).

(A) *Platyamoeba plurinucleolus* with a glycocalyx composed of a dense basal layer and a fuzzy fibrous region.

(B) *Vannella devonica* with distinct glycostyles.

These micrographs show obvious differences in the cell surface structure. However, they fail to indicate the inherent variation in the structure of the glycocalyx over a wide range of *Vannella* and *Platyamoeba* isolates. There is considerable variation in the thickness and composition of the glycocalyx among both *Vannella* and *Platyamoeba* species, and within isolates of the same species (Page, 1983; Butler, 1994). Recent studies have demonstrated that the glycostyles may range in size from 40-500 nm (Butler, 1994; Armstrong, unpublished data). It has been shown that the delicate glycostyles can be easily removed from some strains during chemical fixation during TEM preparation (Page & Blakey, 1979). Furthermore, the cell surface structure has not been examined for many *Vannella* and *Platyamoeba* isolates, and none of the morphologically similar *Clydonella* isolates. Finally, it should be noted that glycostyles are not restricted to the *Vannella* genus: glycostyles of different types have been found in other unrelated genera such as *Vexillifera* and *Paramoeba* (Page, 1979b; Cann & Page, 1982). Therefore, the cell surface structure, considered as the most rigid diagnostic feature for the differentiation and classification of fan-shaped isolates, has some obvious limitations. In the absence of rigid morphological and ultrastructural diagnostic features for identification, the natural relationships among these similar organisms remains undetermined.

To complement morphological and ultrastructural examinations, protein and DNA analyses have been extensively used in an attempt to resolve taxonomic uncertainties and to examine phylogenetic relationships among other organisms (eg. Byers *et al.*, 1983; Costas & Griffiths, 1984; Clark *et al.*, 1989; Johnston *et al.*, 1990; Clark & Diamond, 1991a; Hardys *et al.*, 1992).

### **Protein analysis**

To examine protein differences among similar organisms, proteins from crude lysates are separated according to differences in their size using SDS polyacrylamide gel electrophoresis, isoelectric point using a gel with a pH gradient or by conformation with starch or cellulose gels (Nerad & Daggett, 1979; De Jonckheere, 1983; Pernin *et al.*, 1985; Moss *et al.*, 1988). Typically, the total protein composition can be revealed by staining with Coomassie blue. Alternatively, specific proteins such as alcohol dehydrogenase, leucine amino peptidase, malate dehydrogenase, acid phosphatase, glucose phosphate isomerase, amino peptidase and phosphoglucomutase may be detected in the gel matrix after separation (Nerad & Daggett, 1979; Adams *et al.*, 1989; Leonardi *et al.*, 1995). Intra and interspecific

differences in the protein patterns can then be used for identification purposes and to infer phylogenetic relationships.

With respect to amoebae, protein analysis has been almost exclusively performed on axenic cultures from two pathogenic genera; *Acanthamoeba* from the subclass Gymnamoebia, and the amoeboid-flagellate genus *Naegleria*. Analyses of protein patterns or zymograms from *Acanthamoeba* strains have sometimes revealed intraspecific differences and usually have demonstrated interspecific differences. Comparisons of zymogram patterns have shown considerable overlap between species and suggest groupings of *Acanthamoeba* which are not consistent with previous species assignments based on morphological criteria (Costas & Griffiths, 1980, 1984, 1985; De Jonckheere, 1983). In the case of *Naegleria*, protein analysis has been useful for distinguishing of non-pathogenic and pathogenic species (Nerad & Daggett, 1979; Pernin *et al.*, 1985). Although members of the species complex *Naegleria fowleri* produced similar zymograms (Moss *et al.*, 1988), the differences between most species were too extensive for relationships to be elucidated. This supports the view that the times of divergence within this genus are extremely ancient (Adams *et al.*, 1989). Other protein analyses performed on amoebae have surveyed differences among *Entamoeba* species, identified intraspecific variation among strains of *Mayorella* from different habitats, and seasonal variability in protein concentration (Sargeant *et al.*, 1982; Leonardi *et al.*, 1991, 1995).

Protein electrophoresis is a useful tool for assigning isolates to a species when applied in conjunction with morphological characters, but usually provides little indication of phylogenetic relationships. Protein analysis is usually restricted to axenic cultures, but in some instances it is difficult to obtain sufficient cells for protein analysis (Costas & Griffiths, 1984). Moreover, care must be taken to match culture conditions and to ensure that the fitness of the strains under examination are comparable, as these factors can affect protein levels and activities (Leonardi *et al.*, 1995).

### **Analysis of repetitive DNA**

There are three major types of repetitive DNA; mitochondrial DNA (mtDNA), ribosomal DNA (rDNA) and families of highly repetitive non-coding DNA sequences. The simplest way to identify polymorphisms in repetitive DNA between similar organisms is to digest genomic DNA with restriction enzymes. In most instances the

background DNA smear prevents the examination of specific bands of repetitive DNA. However, the examination of DNA from *Acanthamoeba* and *Naegleria* species has demonstrated that several prominent bands are easily detectable over the genomic background smear (McLaughlin *et al.*, 1988). Analysis of *Naegleria* species identified significant interspecific restriction polymorphisms. Strains of *N. fowleri* and *N. gruberi* were quite homogeneous, although there was some variation related to the geographical distribution of the strains (De Jonckheere, 1987, 1988). Restriction analysis of genomic DNA from morphologically identical *Acanthamoeba* strains produced banding patterns comparable to restriction analyses from purified mitochondrial DNA (Kilvington *et al.*, 1991). Restriction analysis of mtDNA demonstrated that there were relatively few fragments of the same size among strains thought to be of the same species, but there was an indication that one restriction fragment is particularly associated with keratitis (Byers *et al.*, 1983; Kilvington *et al.*, 1991).

In other instances, repetitive DNA has been identified from digested genomic DNA using probes. Probes targeted to specific families of highly repetitive DNA, more commonly known as DNA fingerprinting, have been used extensively in forensic science to identify individual humans (Jeffreys *et al.*, 1985). To examine interspecific relationships, probes have been targeted to mtDNA and rDNA to characterise fungi from the *Sclerotinia* genus. Southern hybridization of these probes to the digested genomic DNA demonstrated extensive polymorphisms for the mtDNA, but the interspecific polymorphisms for the rDNA suggested that rDNA may be a useful taxonomic marker for identification of *Sclerotinia* species (Kohn *et al.*, 1988). The use of hybridising rDNA probes to digested genomic DNA, commonly known as ribotyping, has also been used to evaluate the evolutionary divergence within the *Naegleria* genus and has challenged the taxonomic criteria of the pathogenic fungal genus *Rhizoctonia* (Clark *et al.*, 1989; Vilgalys & Gonzalez, 1990).

Over the last six years, a new PCR-based technique has been developed to sample repetitive DNA for polymorphisms. Random amplified polymorphic DNA (RAPD) uses a series of random oligonucleotides to amplify DNA fragments from genomic DNA using PCR (Williams *et al.*, 1990; Welsh & McClelland, 1990). It has been used to establish taxonomic identity, to assess kinship relationships, to analyse mixed genome samples and to create specific probes (Hardys *et al.*, 1992). Comparisons with other techniques demonstrate that RAPD detects more polymorphisms than



restriction analysis of mtDNA and rDNA which in turn identify more polymorphisms than protein analysis (Foolad *et al.*, 1993).

### **Specific analysis of amplified rDNA**

In the last ten years, the use of PCR-amplified rDNA for identification and phylogenetic analysis has gained considerable momentum. The ribosomal genes or the intervening regions of the rDNA repeat may be amplified using primers complementary to flanking sequence elements. In particular, sequence and restriction analysis of amplified ssrRNA genes has been used for identification, phylogenetic analysis and as a means of estimating genetic divergence (eg. Sogin, 1989; Clark & Diamond, 1991a).

Restriction analysis of the ssrRNA gene, known as riboprinting, has been extensively used to examine the relationships between *Entamoeba* species, *Vahlkampfia* species, *Naegleria* species, Anura parasites assigned to the *Trypanosoma* genus and red algae from the *Gracilaria* and *Gracilariopsis* genera (Clark & Diamond, 1991a, 1991b; Schofield *et al.*, 1991; Brown & De Jonckheere, 1994; De Jonckheere, 1994a; Clark *et al.*, 1995). Riboprinting has lead to several taxonomic revisions of the *Naegleria* genus (Clark *et al.*, 1989; De Jonckheere & Brown, 1995). Restriction analysis of the 28S rRNA gene has also been used for examining some species when riboprinting of the ssrRNA gene is complicated by the presence of introns in the ssrRNA gene (De Jonckheere, 1994a).

The usefulness of riboprinting is very much dependent on the sequence variation among the strains to be examined. In some organisms, such as most plant and animal genera, there is very little sequence variation among the ssrRNA genes. In these cases riboprinting is unlikely to detect any polymorphisms between closely related species, although it may be useful for examining relationships among higher taxonomic levels. When there is little variation apparent in the ssrRNA gene, restriction analysis of more variable regions of the rDNA repeat, such as the internal transcribed spacer, may be useful (Anderson & Baker, 1993; Erland *et al.*, 1994; Goggin, 1994). Although riboprinting has application in identification, only a small proportion of the gene is examined and phylogenies are susceptible to random error, and the method of inferring phylogenies from restriction data can be misleading (Swofford *et al.*, 1996).

For a more accurate phylogeny for closely related organisms, the complete ribosomal gene sequences can be compared (Sogin & Elwood, 1986; Bird *et al.*, 1992; Medlin *et al.*, 1994). However, this is technically demanding for even small numbers of isolates, so partial ribosomal gene sequences, particularly of the *ssrRNA* gene, have been used to examine phylogenies where there is significant sequence variation among closely related species (Johnson *et al.*, 1988, 1990; Barta *et al.*, 1991; Rowan & Powers, 1992; Wada *et al.*, 1992; Vossbrinck *et al.*, 1993; De Jonckheere, 1994b; Marché *et al.*, 1995; Maslov *et al.*, 1996).

The *ssrRNA* of prokaryotic and eukaryotic organisms is composed of conserved regions, semi-conserved regions and variable regions. A majority of the sequence variation between closely related organisms is apparent among the variable regions. Nine variable regions have been identified in prokaryotic and eukaryotic *ssrRNA* genes which are denoted as V1-V9, although V6 is more conserved among eukaryotic organisms (De Rijk *et al.*, 1992). Partial *ssrRNA* sequence analysis incorporates one or more of these variable regions to provide a sufficient number of phylogenetically informative sites for the examination of relationships between similar organisms, although no one variable region is consistently the most phylogenetically informative (compare Johnson *et al.*, 1990; Rowan & Powers, 1992; James *et al.*, 1994; Marché *et al.*, 1995).

There are several examples of how partial *ssrRNA* sequence analysis has complemented and contradicted both morphological and protein data to reveal the true relationships between closely related organisms. Species of the amoeba genus *Acanthamoeba*, are morphologically very similar, with only minor differences in cyst morphology used to discriminate between species (Pussard & Pons, 1977). Protein analysis indicated that the taxonomic features were unreliable (Costas & Griffiths, 1980, 1984; De Jonckheere, 1983). However, partial sequence analysis of the *ssrRNA* V3, V5, V7 and V8 regions identified significant sequence variation and a phylogeny which generally supported the use of cyst morphology for classification of species within this genera. The significant sequence variation also suggests that the rate of evolution of the morphological characters has been very slow (Johnson *et al.*, 1990). Partial sequence analysis of *Naegleria* strains also supports morphological studies. Examination of the V2, V3 and V4 regions identified considerable sequence variation between strains assigned to *N. gruberi*, which confirm reports on cyst morphology and protein analysis that *N. gruberi* is a species complex. The analysis

also supported the view that minor taxonomic revisions should be introduced, and new species names should be proposed for *N. italica* sub-species (De Jonckheere, 1994b). Taxonomic studies of symbiotic dinoflagellates have been continually hindered by the lack of informative morphological features and difficulties encountered culturing cells *in vitro*. There is significant sequence variation for the V2 and V4 regions, but in this case the molecular phylogeny suggests that the morphological characters traditionally used in dinoflagellate classification may misinterpret phylogenetic relationships (Rowan & Powers, 1992).

In other instances there may not be sufficient sequence variation to examine the phylogenetic relationships among closely related organisms. For example, examination of partial sequences from parasitic platyhelminths and yeast suggest that partial *ssrRNA* sequences are more useful for resolving relationships among families rather than intergeneric or interspecific relationships (Johnston *et al.*, 1993; Lumb *et al.*, 1993; James *et al.*, 1994).

### **Use of ribosomal RNA probes**

Sequence comparisons of *ssrRNA* genes have also been used to identify unique sequence primers in PCR-based assays for specific identification of prokaryotes and eukaryotes (Olsen *et al.*, 1986; Amman *et al.*, 1991; Vodkin *et al.*, 1992; Bently *et al.*, 1993; Marché *et al.*, 1995). Ribosomal RNA probes, labelled with fluorochromes or haptens, have also been extensively used to identify whole bacterial and archaeobacterial cells from a variety of environments including activated sludge, soil, wheat rhizosphere and from within eukaryotic cells (Amman *et al.*, 1990; Zarda *et al.*, 1991; Embley *et al.*, 1992; Hahn *et al.*, 1992; Wagner *et al.*, 1994). This technology has been adapted to eukaryotes for the identification of metabolically active protists (Lim *et al.*, 1993), and specific groups such as *Cryptosporidium* (Johnson *et al.*, 1993) and naked amoebae (Stacey, unpublished).

### **Summary of chapter**

Characterisation of *ssrRNA* genes by restriction analysis with just two enzymes has demonstrated that there is a significant amount of sequence variation among the morphologically similar fan-shaped amoebae (Chapter 2). In this chapter, partial *ssrRNA* gene sequence comparisons from *Vannella*, *Platyamoeba* and morphologically similar amoebae isolated from the field were examined. In conjunction with light microscopical observations (Chapter 2), comparisons of



sequence data and ultrastructural features were made to identify unknown field isolates. Sequence comparisons were also used to examine the genetic diversity of named and unnamed organisms, and to identify potential sequence elements which may be useful for the development of species or genera specific probes for identification of fixed cells by hybridization. Phylogenetic inference programs were used to examine the relationships among these morphologically similar isolates, and the usefulness of taxonomic characters, such as the cell surface structure, as indicators of phylogenetic relationships was also assessed.

## Materials and methods

### **Amplification of partial ssrRNA genes**

Examination of previous partial sequencing reports and comparisons of the *Vannella anglica* ssrRNA sequence with other aligned ssrRNAs suggests that most sequence variation is present in the V1, V2, V4, V7 and V8 regions. Thus, comparative sequence analysis was used to design primers for the PCR amplification of the V1+V2, V3, V4 and V7+V8 regions. The primer sequences, positions relative to the *V. anglica* sequence (Chapter 3), and the variable regions which they amplify are shown in Table 5.3. To assess the efficiency of the primer sets, an initial survey was performed using a set of eight genomic DNA preparations from *Vannella* and *Platyamoeba* isolates. From these amplification results, the F6R7 primer set was chosen to amplify 18 *Vannella*, *Platyamoeba* and unidentified isolates for cloning and partial sequencing of the ssrRNA V7 V8 region.

Amplifications were performed in 500 µl Eppendorf tubes containing 2 µl of genomic DNA (typically 10-100 ng), 5 µl of 10x PCR buffer (200 mM Tris-HCl pH 8.4, 500 mM potassium chloride), 0.05% w/w W-1 detergent, 80 pmol of each primer, 25 µM of each ultra-pure dNTP, 2.5 mM magnesium chloride and sterile Millipore water to a total volume of 48 µl. The reaction mixture was overlaid with 100 µl of light mineral oil and preheated to 95°C for 5 min, cooled to 72°C and 2 µl (2.5 units) of *Taq* DNA polymerase (Gibco BRL) added. The PCR reaction mixture was then subjected to 32 cycles of denaturing at 94°C for 80 sec, annealing at 60°C for 80 sec and elongation at 74°C for 2 min, followed by a final 5 min elongation on a Minicycler Model PTC-150 (MJ Research). The ramping rate for the thermal cycler was maintained at 1°C every 2 secs for changes between each step.

### **Cloning the ssrRNA V7 V8 variable region**

The V7 V8 ssrRNA gene fragments from 18 *Vannella*, *Platyamoeba* and unidentified isolates were purified, and cloned using a TA cloning kit (Invitrogen). To purify the V7 V8 ssrRNA gene fragments, the PCR products were separated on a low melting point agarose gel, the bands of the appropriate size were excised from the gel, cleaned using GeneClean II (BIO101) and resuspended in 20 µl of TE buffer. Six microlitres of the DNA solution was ligated with 50 ng of linearised pCRII TA-

**Table 5.3** Primers for the PCR-amplification of partial *ssrRNA* gene sequences. All forward primers are given for the non-coding strand in a 5' to 3' direction. All reverse primers are given for the coding strand in a 5' to 3' direction. All positions are relative to the *V. anglica* *ssrRNA* sequence from Chapter 3.

	Forward Primer		name	position	Reverse Primer		name	sequence	position	Variable Region
	sequence									
F1	CTGGTTGATCCTGCCAG			3-19			R1	TCTCCGGAATCGAACCC	399-383	V1+V2
F2	AGGGTTCGATTCCGGAG			382-398			R2	GGAATTACCGCGGCTGG	598-583	V3
F3	CGGTAATTCCAGCTCCAA			589-606			R3	CCAAGAATTTCACCTCTG	1041-1024	V4
F3a	AGGTCTGGTGCCAGCAGC			569-586			R3a	ATCCAWGMATTTTCACC	1043-1028	V4
F6	TGGTGCATGGCCGTTC			1413-1428			R7	CGGGGAACATGTGTGGC	1804-1788	V7+V8

cloning vector with 4 units of DNA ligase in a total volume of 10  $\mu$ l for 4 hours at 14°C. Half of the ligation mixture was used to transform a 50  $\mu$ l aliquot of competent One Shot cells by incubating on ice for 30 min, heat-shocking at 42°C for 30 sec and returning to ice for a further 2 min. The transformation mixture was made up to 0.5 ml with SOC media (Appendix I) and incubated at 37°C for 1 hour. To select for transformants, aliquots were plated out onto nutrient agar media containing 100  $\mu$ g/ml ampicillin overlaid with 40  $\mu$ l of 20 mg/ml X-Gal, incubated overnight at 37°C and examined for the presence of white, recombinant colonies and blue non-recombinant colonies. The competency of the cells was assessed by determining the transformation frequency of pUC18, and a DNA-free control ligation was also included to ascertain the religation frequency of plasmid pCRII. To determine whether or not the white colonies contain recombinant plasmids with the *ssrRNA* gene fragment, *EcoRI* sites which flank the TA cloning site were utilised. Plasmid DNA was isolated, digested with *EcoRI* to excise the insert, which was then analysed by agarose gel electrophoresis.

### Sequencing cloned gene fragments

Of the 18 cloned *Vannella*, *Platyamoeba* and unidentified V7 V8 *ssrRNA* gene, five *Vannella* isolates, four *Platyamoeba* isolates and three unidentified isolates were sequenced. A listing of the 12 sequenced isolates, along with short identifications, strain numbers and some taxonomic characters are given in Table 5.4. To estimate the rate of misincorporation of incorrect nucleotides by *Taq* DNA polymerase during PCR-amplification, and to identify the correct nucleotide at any ambiguous sites, multiple clones were sequenced for each isolate. Initially two clones were sequenced for each isolate, but, if an ambiguity was apparent between the two clones, a third clone was sequenced. Recombinant plasmids containing the V7 V8 *ssrRNA* region were denatured, washed, dried and sequenced using Sequenase version 2.0 (Amersham Life Sciences) with M13 forward and reverse primers as described in Chapter 3.

### Primary sequence comparisons

Sequences were aligned using the *V. anglica* *ssrRNA* primary and secondary structure (Chapter 3) with the multiple sequence editor ESEE (Cabot & Beckenbach, 1989). Gaps were introduced into the sequences to aid alignments of secondary structure features. The alignment was analysed for sequence elements, which may be used for directing ribosomal probes for identification purposes. The

**Table 5.4** Listing of isolates used in partial *ssrRNA* sequence analyses.

Short Id.	Isolate	Strain	Habitat	Greatest dimension	L:B ratio	Floating form pseudopodia	Cysts	Nucleolar material	Surface structure
Vann.angli	<i>Vannella anglica</i>	1589/8	M	21-24 µm	1.0-1.2	S-M	Absent	Central	Glycostyles
Vann.simpl	<i>V. simplex</i>	1589/3	F/W	42-52 µm	0.8	L	Absent	Central	Glycostyles
Vann.arabi	<i>V. arabica</i>	1589/7	M	27 µm	0.8	S	Absent	Central	Glycostyles
Vann.septr	<i>V. septentrionalis</i>	1589/10	M	21-23 µm	0.8-0.9	L	Absent	Central	Glycostyles
Vann.lata	<i>V. lata</i>	1589/12	F/W	33 µm	0.6	L	Absent	Central	Glycostyles
Plat.maine	<i>Platymoebe mainensis</i>	1565/1	M	26 µm	1.1	M	Absent	Central	Glycocalyx
Plat.placi	<i>P. placida</i>	1565/2	F/W	25 µm	1.2	S-M	Present	Central	Glycocalyx
Plat.austr	<i>P. australis</i>	1565/9	M	32 µm	1.0	L	Absent	Central	Glycocalyx
Plat.plu11	<i>P. plurinucleolus</i>	1565/11	M	11-26 µm	1.0-1.1	M-L	Absent	Parietal	Glycocalyx
VP3	VP3	VP3	M	22 µm	0.8	M-L	Absent	Unknown	Unknown
Species E	Species E	VP14	M	10 µm	1.5	S-M	Absent	Unknown	Unknown
SIA	SIA	VP31	F/W	25 µm	1.1	S-M	Present	Unknown	Unknown

With respect to habitat, M = marine and F/W = freshwater and for the floating form pseudopodia, S = short, M = medium and L = long. The values for the greatest dimensions and length: breadth ratios are averages.

potential use of ribosomal probes was assessed using the CHECK\_PROBE program from the ribosomal RNA database project (Maidak *et al.*, 1996).

### **Phylogenetic analysis of V7 V8 DNA sequences**

To determine the phylogeny of the 12 morphologically similar naked, fan-shaped amoebae, the aligned sequences were subjected to parsimony and maximum likelihood analyses. Parsimony analyses were performed using PAUP software (Swofford, 1993). The most parsimonious trees were found using a branch and bound tree searching algorithm (Hendy & Penny, 1982). Data sets were resampled 100 times using the bootstrap technique (Felsenstein, 1985) to indicate the support for the topology of the trees. Maximum likelihood analyses were performed using FASTDNAML (Olsen *et al.*, 1994), with options for the randomization of input sequences, and the global rearrangement of branches during the search for the most likely tree.

To root the phylogenetic trees, the partial *ssrRNA* sequence from a recent common ancestor was introduced to the analyses, and used as an outgroup. Phylogenetic analyses using the complete *ssrRNA* gene from *V. anglica* previously demonstrated that *E. histolytica* was the closest unambiguous ancestor to the flat fan-shaped amoebae, for which *ssrRNA* sequence data is available (Chapter 4), so this taxon was used as an outgroup to root the tree.

To assess the usefulness of diagnostic features used to differentiate isolates among the naked, fan-shaped amoebae, a series of constraint trees were constructed using MACCLADE (Maddison & Maddison, 1992). Each constraint tree split the isolates into two or three groups on the basis of a morphological or ultrastructural feature. Using these constraints, the most parsimonious tree was determined, using a branch and bound searching algorithm, for each diagnostic feature. The usefulness of these features was assessed by comparing the length of each constraint tree with the most parsimonious unconstrained phylogenetic tree. Groups for the constraint trees were derived from data presented in Table 5.4. To assess the usefulness of the cell surface structure, the character which is currently used to classify fan-shaped amoebae into the genera *Vannella* and *Platyamoeba*, the isolates were split into two groups with all the *Vannella* isolates with glycostyles in one group and all *Platyamoeba* with less differentiated cell surfaces in the other. Isolates VP3, Species E and SIA were left unconstrained for this feature. For the greatest dimension, the

isolates were split into three groups, with isolates with an average greatest dimension up to 15  $\mu\text{m}$  in one group, isolates between 16-30  $\mu\text{m}$  in a second, and isolates 31  $\mu\text{m}$  and above in the third. For the length:breadth ratio, the isolates were split into two groups, with isolates with a ratio less than one in one group and isolates with a ratio of one and above in the second. Finally, for the length of the floating form pseudopodia, the isolates were split into three groups with isolates possessing short or short to medium length pseudopodia in one group, isolates with medium or medium to long pseudopodia in a second group, and the isolates with long pseudopodia in a third group.

### **Transmission electron microscopy**

Electron microscopy was used to further investigate the identity of VP3 and Species E. The ultrastructure of *V. anglica*, which has been detailed previously (Page, 1980a), was also examined as a control of the preparation protocol. All three isolates were rinsed from the surface of agar plates or culture flasks with sterile seawater, transferred to a centrifuge tube and centrifuged for 5 min at 2500 rpm. The supernatant was removed and the cells were resuspended in 5 ml of buffered fixative (0.05 M cacodylate buffer pH 7.2, 0.6 M sodium chloride containing 2.5% v/v glutaraldehyde) for 30 min. The cells were centrifuged at 2500 rpm for 5 min, the supernatant was removed and the cells were rinsed twice in 1 ml of 0.05 M cacodylate buffer. The cells were then centrifuged, the supernatant was removed and the cells were post fixed in 2% osmium tetroxide in 0.05 M cacodylate (pH 7.2) for 1 hour at 4°C. Following centrifugation, the supernatant was removed and the cells were resuspended in fresh buffered osmium for 30 min. The cells were rinsed in distilled water twice, resuspended in approximately 200  $\mu\text{l}$  of 0.8% molten agar at 40°C and centrifuged at 2500 rpm for 3 min. The agar was set by plunging the centrifuge tube into ice-cold water for 2 min. The agar block containing cells was dissected out, carefully sliced into small pieces and transferred to a universal tube. The agar pieces were then dehydrated with consecutive 10 min, 1 ml washes of 30%, 50%, 70%, 80%, 95% and 100% acetone solutions followed by a final 15 min wash with 1 ml of dry 100% acetone. The acetone was pipetted off and the blocks were infiltrated by placing in a mixture of 3 ml of dry 100% acetone and 3 ml of Spurr's epoxy resin (Spurr, 1969). The mixture was left overnight in a fume cupboard with the tube open to allow the acetone to evaporate. The resin mixture was replaced with fresh resin for a further 12 hours and then the agar pieces were transferred to beam capsules containing fresh resin and polymerised overnight at 70°C.

The resin blocks were trimmed and ultrathin silver and gold sections were cut with a diamond knife. The sections were picked up on copper grids, stained with freshly prepared 3% uranyl acetate in 30% ethanol for 1 min and freshly prepared 0.2 M lead nitrate, 0.2 M sodium citrate for 2 min (Reynolds, 1963), rinsed in distilled water and dried. Stained sections were examined in a JOEL JEM-100S Electron Microscope at 60 kv at magnifications between 2,000 and 50,000 times.



## Results

### **Partial ssrRNA gene amplifications**

To assess the usefulness of the PCR primer sets to amplify variable regions of the ssrRNA genes, amplifications were performed upon the plasmid DNA containing the complete *Vannella anglica* ssrRNA and a set of eight *Vannella* and *Platyamoeba* genomic DNA preparations. The amplifications for the primer sets F1R1, F2R2, F3R3 and F6R7 are shown in Figure 5.2 A, B, C and D, respectively.

Amplifications were generally successful from the F1R1, F2R2 and F6R7 primer sets. For primer set F1R1, all the genomic DNA preparations produced an amplification product the same size as the *V. anglica* ssrRNA, at approximately 410 bp. For the primer set F2R2, all the genomic DNA preparations except *V. platypodia* produced an amplification product of approximately 220 bp, again the same size as *V. anglica* control amplification. The unsuccessful amplification from *V. platypodia* suggests that sufficient mismatches are apparent in the template sequence to prevent primer annealing under these particular reaction conditions. Amplification of the V7 and V8 regions using the F6R7 primer set was successful. The positive control, and the complete set of eight genomic DNA preparations all produced an amplification product of approximately 450 bps. The estimated sizes of 410, 220, and 400 bp, determined by agarose gel electrophoresis for the F1R1, F2R2 and F6R7 amplifications, were very similar to the expected sizes of 396, 217 and 392 bp determined from the complete sequence of *V. anglica* in Chapter 3. Moreover, for each set of primers no amplifications were apparent from the DNA-free negative control reactions. This would suggest that the amplification products produced were indeed the partial ssrRNA gene fragments required, and additional confirmatory hybridization experiments were considered unnecessary.

Amplification of the V4 variable region with the F3R3 primer set was completely unsuccessful. These primers were designed to anneal to regions which are conserved in an alignment of ssrRNA sequences from a variety of eukaryotic taxa, to amplify a fragment of approximately 450 bp. However, these primers could not amplify the gene fragment from any of the genomic DNA preparations or even the plasmid DNA containing the complete *V. anglica* ssrRNA gene. The inability to amplify the V4 region may have been due either mismatches between the primers and the template

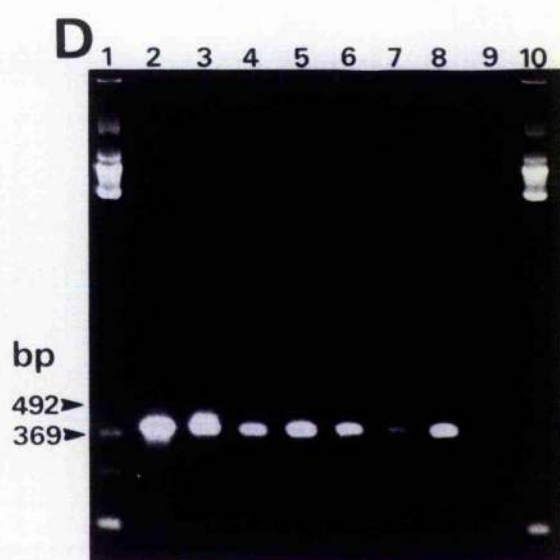
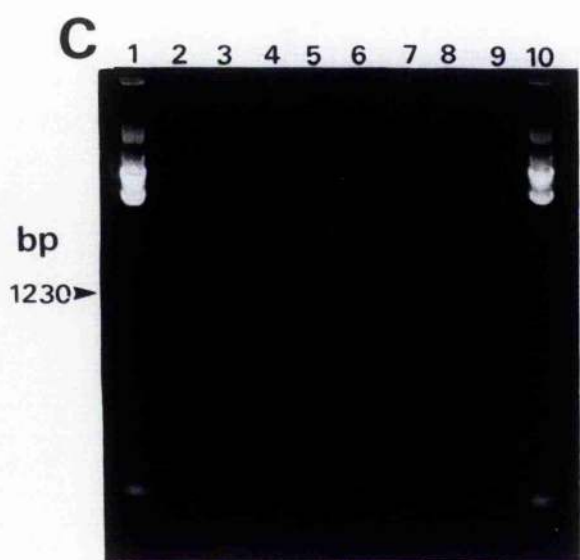
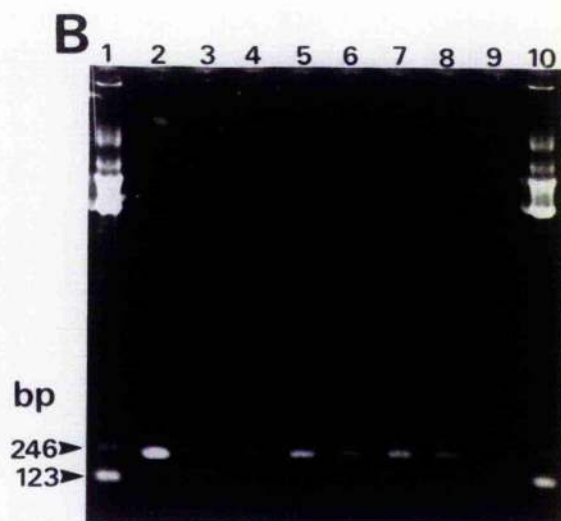
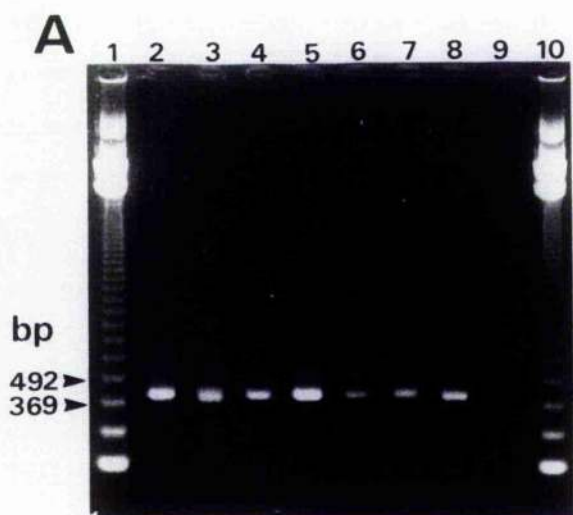
DNA or to secondary structure of the rDNA in this region. The secondary structure may prevent one or both of the primers from annealing to the template DNA or the secondary structure may prematurely terminate the DNA replication process. A single weak product of approximately 1.2 kbp was amplified from *V. platypodia* which suggests that only one of the primers may have been annealing efficiently, as a product of 1.2 kbp is far too large for the V4 region alone. When the *ssrRNA* sequence data from *V. anglica* became available it was clear that there were mismatches between the sequence and both primers. Using the *V. anglica* sequence, an alternative set of primers, F3a and R3a, were synthesized but amplifications of the V4 region of the *ssrRNA* gene with these primers were unsuccessful or poor. This would suggest that the secondary structure formation may prevent successful amplification of this region.

---

**Figure 5.2** Partial *ssrRNA* gene amplifications from selected amoebae.

- (A) Amplification of the V1 V2 region with primers F1 and R1.
- (B) Amplification of the V3 region with primers F2 and R2.
- (C) Amplification of the V4 region with primers F3 and R3.
- (D) Amplification of the V7 V8 region with primers F6 and R7.

Lanes 1: 123 bp DNA marker  
 Lanes 2: Plasmid pCRV01.106 DNA positive control  
 Lanes 3: *Vannella platypodia*  
 Lanes 4: *V. simplex*  
 Lanes 5: *V. septrentionalis*  
 Lanes 6: *V. lata*  
 Lanes 7: *Platyamoeba mainensis*  
 Lanes 8: *P. placida*  
 Lanes 9: DNA-free negative control  
 Lanes 10: 123 bp DNA marker



Considering that the primer sets F1R1 and F6R7 each successfully amplified two variable regions covering a 400 bp segment, both of these primers were deemed as prospective candidates for partial *ssrRNA* sequence analysis. Previous analysis has demonstrated that the V7 V8 region is more likely to provide more phylogenetically informative sites, so this region was chosen. The amplifications for the V7 and V8 regions of the *ssrRNA* gene for *Vannella*, *Platyamoeba* and unknown isolates are shown in Figure 5.3 A, B and C. All the isolates produced a product of approximately 380-400 bp, except Species E, which produced a considerably larger product of about 430 bp. In the case of *V. platypodia*, *P. placida* and *P. stenopodia* weak auxiliary bands were also apparent.

### Cloning

Generally, the transformation frequency into competent One Shot cells was quite low. Plasmid pUC18 transformed at a frequency of  $7.9 \times 10^5 \mu\text{g}^{-1}$ . With the TA-cloning plasmid pCRII, ligations with the V7 V8 *ssrRNA* amplicons transformed with a frequency ranging from  $8.8 \times 10^2$  to  $1.2 \times 10^4 \mu\text{g}^{-1}$ , with an average frequency of  $4.6 \times 10^3 \mu\text{g}^{-1}$ . The transformation frequency for the pCRII religation in the absence of DNA was  $1.0 \times 10^3 \mu\text{g}^{-1}$ ; this was marginally higher than one of the 18 transformation frequencies of pCRII ligated with a DNA fragment, but almost five times lower than the average frequency. Religations probably resulted from the removal of the single 5' thymidylate overhangs and blunt-end ligation.

To screen for potentially recombinant plasmids, blue-white selection was used on media containing X-Gal, where a blue colony indicates a non-recombinant plasmid and a white colony may indicate a recombinant plasmid. For the religation of pCRII, in the absence of a compatible PCR product, the frequency of white colonies as a proportion of the total number of colonies was 3.8%. In the presence of the V7 V8 *ssrRNA* PCR amplifications the proportion of white colonies ranged from 28 to 68%. Plasmid DNA was isolated from five white colonies from each of the 18 transformations involving ligations with *ssrRNA* gene fragments. The plasmid DNA was digested with *EcoRI* and examined by agarose gel electrophoresis. All ninety plasmids released a DNA fragment of approximately 400 bp. These results demonstrate that although the transformation frequency was reasonably low, the ligation of the amplicons into the pCRII plasmid and the blue-white selection for the presence of the recombinant plasmids were efficient.

**Figure 5.3** Partial ssrRNA gene amplifications from the V7 V8 variable region.

**(A)** Amplifications from *Vannella* spp. genomic DNA extractions

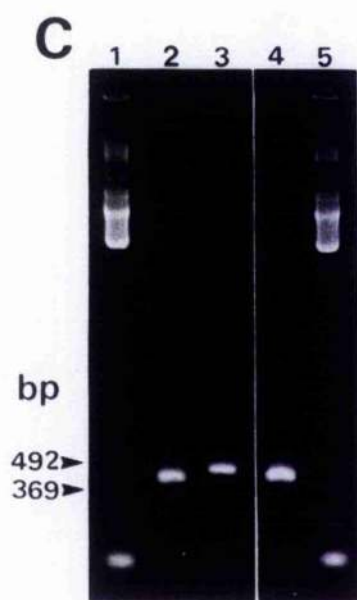
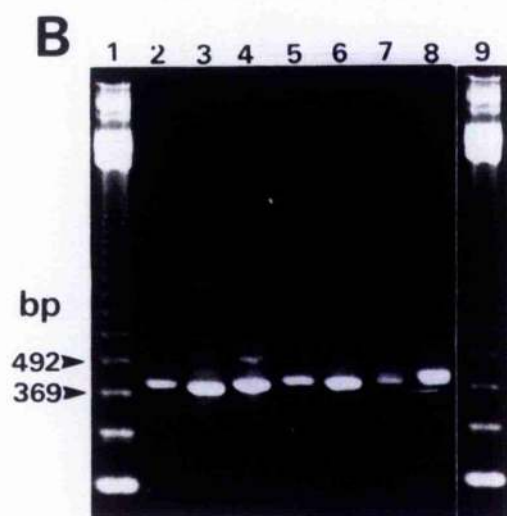
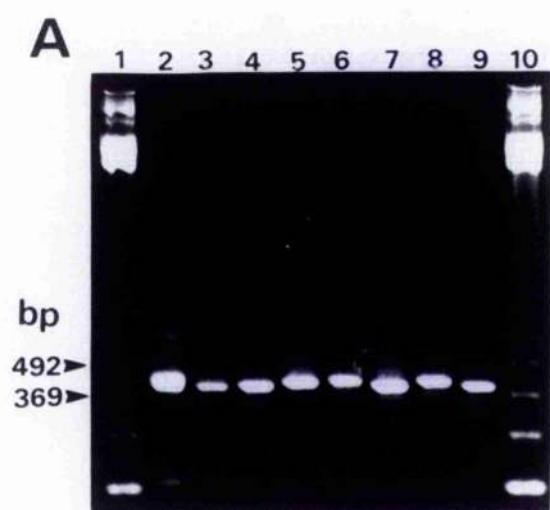
Lane 1: 123 bp ladder  
Lane 2: *V. platypodia*  
Lane 3: *V. simplex*  
Lane 4: *V. devonica*  
Lane 5: *V. arabica*  
Lane 6: *V. anglica* 1589/8  
Lane 7: *V. septrentionalis*  
Lane 8: *V. anglica* 1589/11  
Lane 9: *V. lata*  
Lane 10: 123 bp ladder

**(B)** Amplifications from *Platyamoeba* spp. genomic DNA extractions

Lane 1: 123 bp ladder  
Lane 2: *P. mainensis*  
Lane 3: *P. placida*  
Lane 4: *P. stenopodia*  
Lane 5: *P. calycinucleolus*  
Lane 6: *P. australis*  
Lane 7: *P. bursella*  
Lane 8: *P. plurinucleolus* 1565/11  
Lane 9: 123 bp ladder

**(C)** Amplifications from unidentified isolates genomic DNA extractions

Lane 1: 123 bp ladder  
Lane 2: VP3  
Lane 3: Species E  
Lane 4: SIA  
Lane 5: 123 bp ladder



### Partial ssrRNA V7 V8 sequencing

The partial ssrRNA gene sequence encompassing the V7 and V8 variable regions, from helix 38 to the reverse of helix 32 (positions 1412 to 1803 of the *V. anglica* ssrRNA sequence), was determined from 12 isolates; five *Vannella* spp., four *Platyamoeba* spp., and three unidentified isolates. The sequences for each isolate were determined from two or three clones. One or two ambiguities were apparent for some of the isolates presumably due to misincorporation of nucleotides by *Taq* DNA polymerase during PCR-amplification. The average error rate was less than one misincorporated base per strain over the 400 bp region sequenced. The sequences were aligned using conserved sequence elements and the predicted secondary structures. The alignments of these sequences, the relative positions of the helices and the total length of each sequence are shown in Figure 5.4. Gaps were introduced to aid alignment of primary and secondary structure. The alignment therefore covers a total of 547 positions.

The length of the V7 V8 region ranged from 380 bp in *P. placida* and isolate SIA, to 401 bp in *V. arabica*, excepting Species E which was considerably larger than all the other isolates with a length of 438 bp. The greater length of Species E can be attributed to an extended helix 43, the secondary structure of which could not be determined. The length variation among the remaining isolates was due to the variable size of the terminal loop of helix 43 (positions 170-179), the small loop within helix 43 (positions 193-211) and the small AT-rich expansion region between helices 45 and 46 (positions 352-383).

Comparisons of the sequence data illustrated that helices 38, 39, 40 41, 42 and 47 are generally well conserved with only a few differences between the sequences. There were some differences within helix 48, but a large majority of the variation was constrained within the variable V7 (helices 43 and 44) and V8 (helices 45, 46 and the intervening expansion segment) regions.

For the 442 positions included in the restricted data set, the number of differences between Species E and the other isolates varied from 62 to 77 differences. The number of differences between pair-wise comparisons of the seven isolates of the marine clade generally varied from nine to 16, although the number of differences for pair-wise comparisons involving *P. plurinucleolus* were as high as 25 due to a nine base insert within the loop of helix 43.

**Figure 5.4** Alignment of partial *ssrRNA* V7 V8 sequences. A legend with details of this figure is given at the end of the alignment.

	1	38	39	40	40'	41	50
Vann.angli	TGGTG-	CATGGCCGTT-	CTTAGTTGGTGA-	GTCATTGTC-	--	TGGTT-A	
Vann.simpl	.....	.....	.....	.....	.....	.....	
Vann.arabi	.....	.....	.....	.....	.....	.....	
Vann.septr	.....	.....	.....	.....	.....	.....	
Vann.lata	.....	.....	.....	.....	.....	.....	
Plat.maine	.....	.....	.....	.....	.....	.....	
Plat.placi	.....	.....	.....	.....	.....	.....	
Plat.austr	.....	.....	.....	.....	.....	.....	
Plat.plu11	.....	.....	.....	.....	.....	.....	
VP3	.....	.....	.....	.....	.....	.....	
Species E	.....	.....	.....	.....	.....	.....	
SIA	.....	.....	.....	.....	.....	.....	

	41'	39'	42	43	100
Vann.angli	ATTCCGAT-	AACGAACGAGACCTT-	TA----	CCATACTAA-	ATAGTT-TG
Vann.simpl	.....	.....	.....	.....	-C.....-A
Vann.arabi	.....	.....	.....	.....	.....
Vann.septr	.....	.....	.....	.....	.....
Vann.lata	.....A.....	.....T.....	.....A.....	.....-G.....	.....A--A
Plat.maine	.....	.....	.....	.....	.....
Plat.placi	.....T.....	.....	.....G----	.....-G.....	.....A--
Plat.austr	.....	.....	.....	.....	.....
Plat.plu11	.....	.....	.....	.....	.....
VP3	.....	.....	.....	.....	.....
Species E	.....	.....	.....C-A----	.....	.....-AA
SIA	.....	.....	.....	.....-G.....	.....A--

	111	150
Vann.angli	-ATCTATT-	-----
Vann.simpl	-TAT.TA.	-----
Vann.arabi	-.T.....	-----
Vann.septr	-----	-----
Vann.lata	-T...TA.	-----
Plat.maine	-.T...A-	-----
Plat.placi	-TA..TG.	-----
Plat.austr	.....A-	-----
Plat.plu11	-.T...A-	-----
VP3	.....A-	-----
Species E	-TG.AGACT--	-CTTCTACTCAAGAGCGTCGGTGGGGTGAGCAAGTCTTC
SIA	-TG..TG.	-----

	183	43'	200
Vann.angli	-----	CTTATTGT-	ATAGATT-TAT----
Vann.simpl	-----	AC....	.....A-----
Vann.arabi	-----	CA..TA..A-	.....
Vann.septr	-----	.A.T...A-	.....
Vann.lata	-----	.ACT.CTC-	.....CAA-----
Plat.maine	-----	.A.T..T.	.....
Plat.placi	-----	-AT..TA-	TACAGG.
Plat.austr	-----	.A.T..T.	.....
Plat.plu11	-----	-AAT.C...T.	.....GATA
VP3	-----	.A.T..T.	.....
Species E	GGATGAGCTTATTCTGTCCGGACTC.	GGAT.	---AG.CTG.A-----
SIA	-----	AT..TA-	TACA.G.



**Figure 5.4 Continued.**

[illegible]

	44'	42'	38'	300'
Vann.angli	TA-AACC--AAAGGAAGT--	AAAGGCGATAACAGG	PCTG-TGATGCCC-T	
Vann.simpl	C.-.-.-.-.-	-.-.-.-.-	-.-.-.-.-	-.-.-.-.-
Vann.arabi	-.-.-.-.-	-.-.-.-.-	-.-.-.-.-	-.-.-.-.-
Vann.septr	-.-.-.-.-	-.-.-.-.-	-.-.-.-.-	-.-.-.-.-
Vann.lata	C.-.-.-.-.-	-.-.-.-.-	-.-.-.-.-	-.-.-.-.-
Plat.maine	-.-.-.-.-	-.-.-.-.-	-.-.-.-.-	-.-.-.-.-
Plat.placi	C.-.T.-.-.-.-.-	-.-.-.-.-	-.-.-.-.-	-.-.-.-.-
Plat.austr	-.-.-.-.-	-.-.-.-.-	-.-.-.-.-	-.-.-.-.-
Plat.plull	-.-.-.-.-	-.-.-.-.-	-.-.-.-.-	-.-.-.-.-
VP3	-.-.-.-.-TT--	-.-.-.-.-	-.-.-.-.-	-.-.-.-.-
Species E	-.-.G.A--G.T.-.-.-	-TTG.-.-.A.A.-.-.-	-.-.-.-.-	-.-.-.-.-
SIA	C.-.-.T.-.-.-.-.-	-.-.-.-.-	-.-.-.-.-	-.-.-.-.-

	36'	34'	45	350
Vann.angli	TAG-ATGTTCTGGG-CTGCACGCGGCTACAATGA-TAATCT-CAACGAG			
Vann.simpl	.....		GA-..T.A..	
Vann.arabi	.....		A-.....	
Vann.septr	.....		G-.....	
Vann.lata	.....	A.A.....	C-.....T.A..	
Plat.maine	.....		A-.....	
Plat.placi	.....C.....		T.GT-..T...	
Plat.austr	.....		A..CA.....	
Plat.plu11	.....		G.T-.....	
VP3	.....		A-.....	
Species E	.....		C...-CGGATG-.....	
SIA	.....C.....		C...-T.GT-..T.A..	

	352		383	46	400
	1		1		1
Vann. angli	T-----ATTTATTTTATTATATT-----	GCTTTATT--GG-AA--			
Vann. simpl	-----A..T.ACG..AT-----	-----C-----			
Vann. arabi	-----A.A.T.A.T.A.TATATT-----	-----G...C-----			
Vann. sept	-----AAA.T.A..T..ATC-----	-----C-----			
Vann. lata	-----TA..TCAACT.GT.TGAAATTT-----	A...TCC--T..T-----			
Plat. maine	-----A.....T.....	-----CC-----			
Plat. placi	-----T.CG.AACA.G-----	A..AC.CC--T-----			
Plat. austr	-----TA.AT...AT..TAT-----	CG...CC-----			
Plat. plull	-----A..T..T.TA.ATTTT-----	A...A.CC-----			
VP3	-----AAG.TAACAC..TATC-----	-----AGCC-----			
Species E	-----TA..C-----	A...GGCC--A-T-----			
SIA	-----T...TCAAA-----	A..A..CC--T-----			

**Figure 5.4 Continued.**

	401 1	46'	45'	450 1
Vann.angli	---	AATA--TGGC-TAATCTTA--	AACGATTAT-CGTG--	CTT-GGGATA
Vann.simpl	---	GG.....	GTYY.....	
Vann.arabi	---	GG.....	T.....	
Vann.septr	---	G.....	C.....	
Vann.lata	---	GGA.....T-.....	C.C--..AT.G.-T.....	
Plat.maine	---	GG.....	T.....	
Plat.placi	---	GG.G--..T-.....	C.A.....	
Plat.austr	---	GG.....	T.....	
Plat.plull	---	GG.T--..T-.....	A.C.....	
VP3	---	GGCT--..	T.....	
Species E	---	GG.C--..T-.....	TTG...TCC--..G-	
SIA	---	GG.G--..T-.....	C.A.....	

	47	47'	33'	48 500 1
Vann.angli	GATGATTGCAATT-ATTTATCTTG-	AACGAGGAATTCCTAGTAATC-GTA		
Vann.simpl	.....	C...CA-..C.....		
Vann.arabi	.....	C.....		
Vann.septr	.....	C.....		
Vann.lata	.....T.....	A.....		
Plat.maine	.....	C.....		
Plat.placi	.....	C.....		
Plat.austr	.....	C.....		
Plat.plull	.....	C.....		
VP3	.....	C.....		
Species E	.....G.A-..C..C-..A.....G.-C.			
SIA	.....	C.....		

	48'	32'	547 1
Vann.angli	-TATCATCAGTGTCGGATGATTACGTCCTGCCCCCTGTACACACCG		392
Vann.simpl	-GC.....GC.....		383
Vann.arabi	.....		401
Vann.septr	.....A.....		393
Vann.lata	-AT.....AT.....T.....		390
Plat.maine	.....		392
Plat.placi	-GT.....AC.....		380
Plat.austr	.....		391
Plat.plull	.....		398
VP3	.....A.....		394
Species E	-AG.....CT...T.....		438
SIA	-GT.T.....AC.....		380

The *V. anglica* partial ssrRNA sequence is given on the first row of the alignment. In the succeeding sequences, identical nucleotides are denoted by a dot; only nucleotides which differ are given. Dashes have been used to aid the alignment of homologous positions. The number and relative position of each helix is displayed in bold above the alignment. The alignment position is given at the end of each line and at other appropriate places, and the number of nucleotides in each sequence is shown at the end of the alignment.

For the freshwater isolates, although there were eight differences between *P. placida* and the morphologically indistinguishable isolate SIA, the number of differences between any other pair-wise combination of freshwater isolates was more than 40 differences.

Sequence comparisons revealed that there were no recognisable sequence elements which could be utilised as genera specific probes, although one site could be useful for specific identification of *Vannella* and *Platyamoeba* species. This 16 nucleotide sequence element 5' AAAGGAAGTAAAGGCG 3' corresponded to positions 1561-1576 of the *V. anglica* ssrRNA sequence (Chapter 3) running from the middle of helix 44' to the bulge between helices 42' and 38' in the V7 region. This element was identical for all *Vannella* and *Platyamoeba* species, VP3 and SIA, but not Species E had six mismatches. The potential of a rRNA probe, 5' CGCCUUUACUCCUUU 3', directed against this element, was assessed using CHECK\_PROBE. No ssrRNA sequences were found with less than three mismatches. Comparisons with *Acanthamoeba castellanii*, *Hartmannella vermiformis*, *Dictyostelium discoideum* and *Entamoeba histolytica* ssrRNA sequences demonstrated that they had eight, five, seven and five mismatches, respectively. These results suggest that this probe may be useful for identification of *Vannella* and *Platyamoeba* species. There were some sequence elements which were unique to particular species. Probes could be easily designed to hybridize to sequence elements within the helix 43 expansion segment of Species E. To a lesser extent, the small insertion in helix 43 of *P. plurinucleolus* could be utilised for the design of a species specific probe.

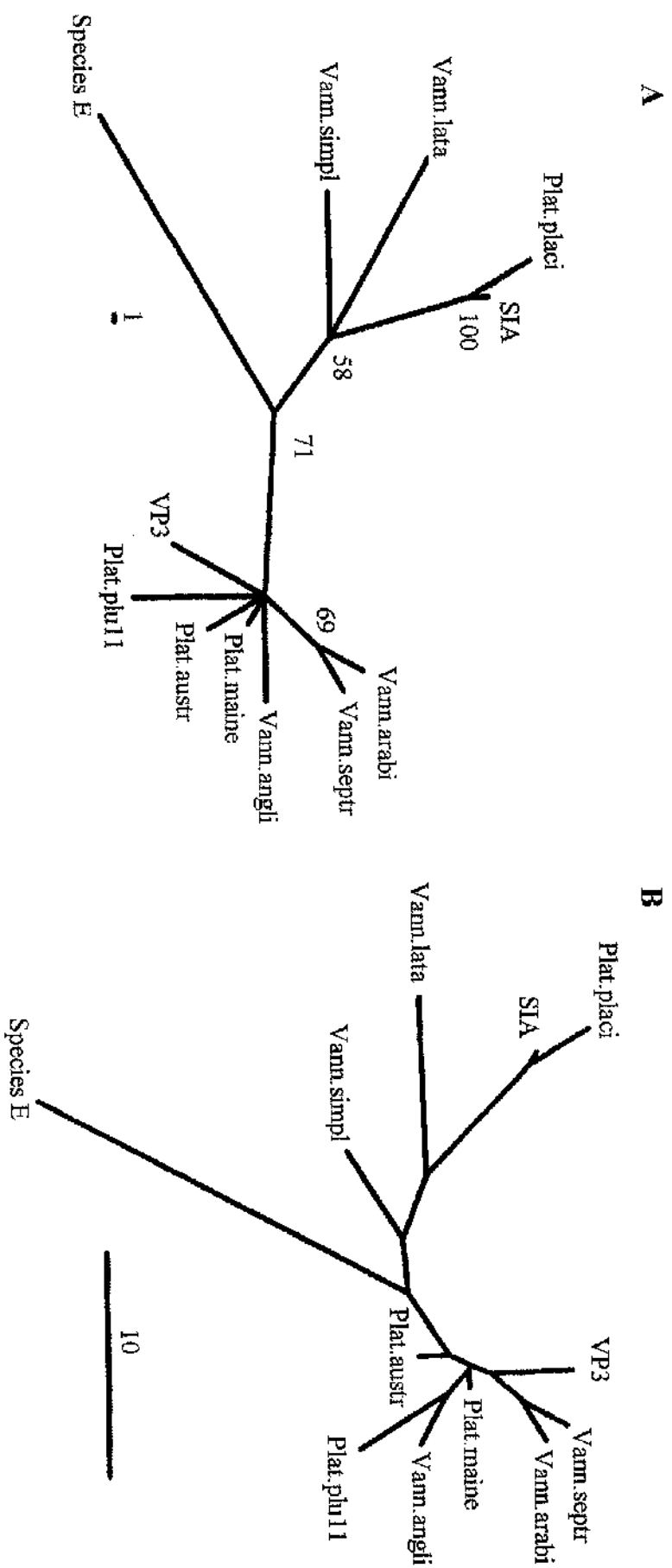
### Phylogenetic analyses on the complete data set

Initial phylogenetic analyses were performed using the complete 542 nucleotide sequence alignment from the 12 sequences using parsimony and maximum likelihood methods. Of the 542 positions, 412 sites were constant, 58 sites are parsimony uninformative due to autapomorphy, leaving 72 informative sites. For the parsimony analysis, a branch and bound tree searching algorithm found 23 similar, most parsimonious trees with a total length of 263 evolutionary changes. The data set was bootstrap resampled 100 times and the unrooted, semi-consensus most parsimonious tree is given in Figure 5.5 A. The parsimony analysis demonstrates that the naked fan-shaped amoebae are composed of two clusters. One cluster, with bootstrap support of 58%, is composed of the freshwater *Vannella* and *Platyamoeba* isolates *V. simplex*, *V. lata* and *P. placida*, as well as the unidentified freshwater

isolate SIA. Within this cluster there is 100% bootstrap support for SIA and *P. placida*. The second cluster contains all the marine *Vannella* and *Platyamoeba* marine isolates (*V. anglica*, *V. arabica*, *V. septrentionalis*, *P. mainensis*, *P. plurinucleolus* and *P. australis*) plus the unidentified marine isolate, VP3. The branching order of the isolates among this cluster is generally unresolved, although there is weak bootstrap support (69%) for a relationship between *V. arabica* and *V. septrentionalis*. There was also bootstrap support (71%) for the separation of the two clusters by the branching of the unidentified isolate Species E. The variation in the topology of the 23 most parsimonious trees is due to variation in the branching order of the isolates within the marine cluster.

The unrooted phylogram from the maximum likelihood analysis, with a negative ln likelihood of 1962.30 is shown in Figure 5.5 B. The maximum likelihood analysis was congruent with the parsimony analysis; the freshwater and marine clusters were evident, and Species E again clearly did not associate with either cluster. Within the freshwater cluster the relationship between SIA and *P. placida* was detected, *V. simplex* branched next followed by *V. lata*. This branching order was observed among all the most parsimonious trees, although there was no bootstrap support. The branching order of the isolates among the marine cluster was resolved with the maximum likelihood analysis, although no statistical analysis was performed to substantiate the order. *Platyamoeba australis* branches first followed by the bifurcation of *V. septrentionalis*, *V. arabica* and VP3, and *V. anglica*, *P. plurinucleolus* and *P. mainensis*. Examination of the most parsimonious trees demonstrated that *P. australis* was also the first taxon to branch among the marine isolates in the parsimony analysis.

The use of all the 542 sites from these partial sequences, however, is open to criticism as some areas of the alignment could not be unambiguously aligned. These areas included the expansion region of Species E and the terminal loop of helix 43, and the small expansion segment between helices 45 and 46. In all subsequent parsimony and maximum likelihood analyses these areas (positions 111-183 and 352-383) were excluded, leaving 442 characters of which 337 were constant, 49 were autapomorphic and 56 were parsimony informative.



**Figure 5.5** Unrooted phylograms of naked, fan-shaped amoebae using sequence data from the complete V7 V8 variable region. (A) Majority-consensus most parsimonious phylogram. The data set was resampled 100 times with the bootstrap method and the most parsimonious trees were determined using a branch and bound algorithm. Bootstrap values less than 50%, and branches occurring in less than 50% of the trees are not shown. The horizontal scale represents one observed nucleotide change. (B) Maximum likelihood phylogram. The horizontal scale represents 10 expected evolutionary changes every 100 nucleotide positions.

### Phylogenetic analyses on a restricted data set

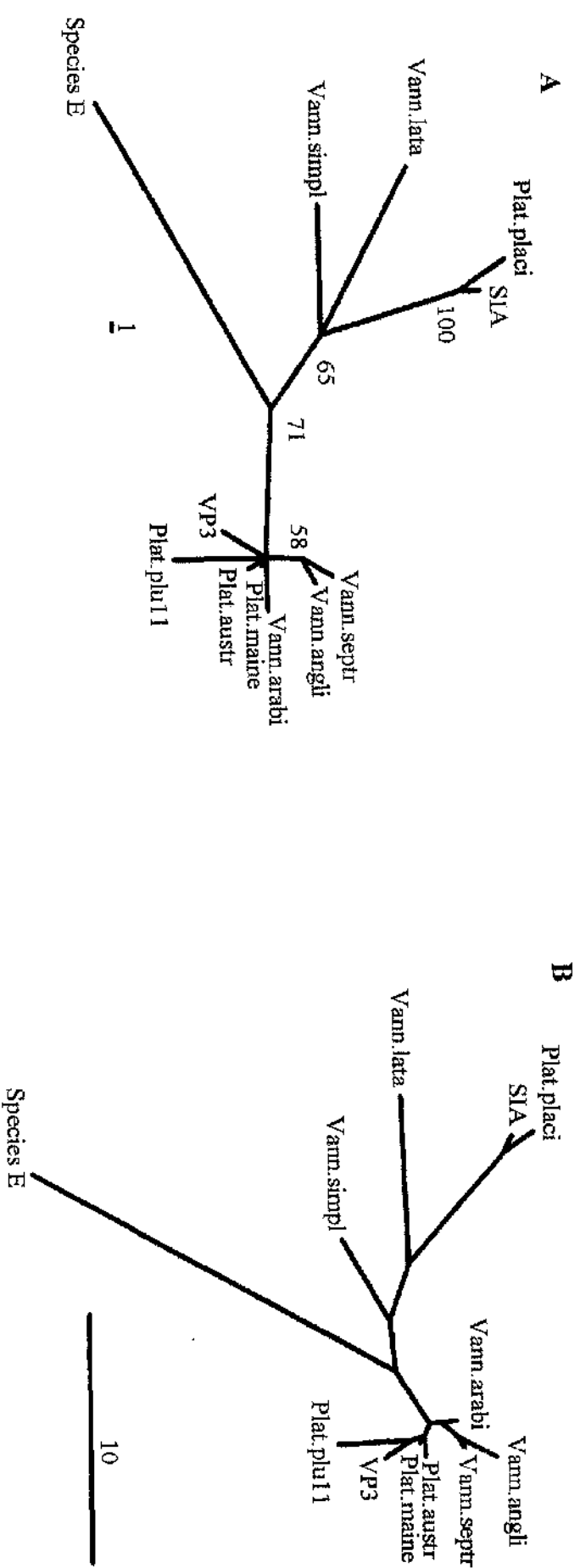
The restricted phylogenetic analyses were not significantly different from the previous analyses. For the parsimony analysis, six most parsimonious trees were found with a branch length of 183. The majority-consensus most parsimonious tree with bootstrap values is shown in Figure 5.6 A. The freshwater and marine clusters were again separated by Species E. The bootstrap support for the freshwater cluster increased to 65%, and within this cluster the 100% support continued for association between SIA and *P. placida*. Within the marine cluster, the branching order was not resolved any further. The weak bootstrap support for *V. arabica* and *V. septrentionalis* in the previous parsimony analysis, was replaced by weak support for a relationship between *V. anglica* and *V. septrentionalis*, with the restricted data set. The bootstrap support for the branch position of Species E remained at 71%.

For the restricted data set, the maximum likelihood analysis produced a tree with a negative ln likelihood of 1438.86 shown in Figure 5.6 B. Like the parsimony analysis, the maximum likelihood analysis only varied from the analysis on the complete data set in the branching order of the isolates among the marine cluster. The *Vannella* isolates formed one clade, and the *Platyamoeba* isolates and VP3 formed a second clade.

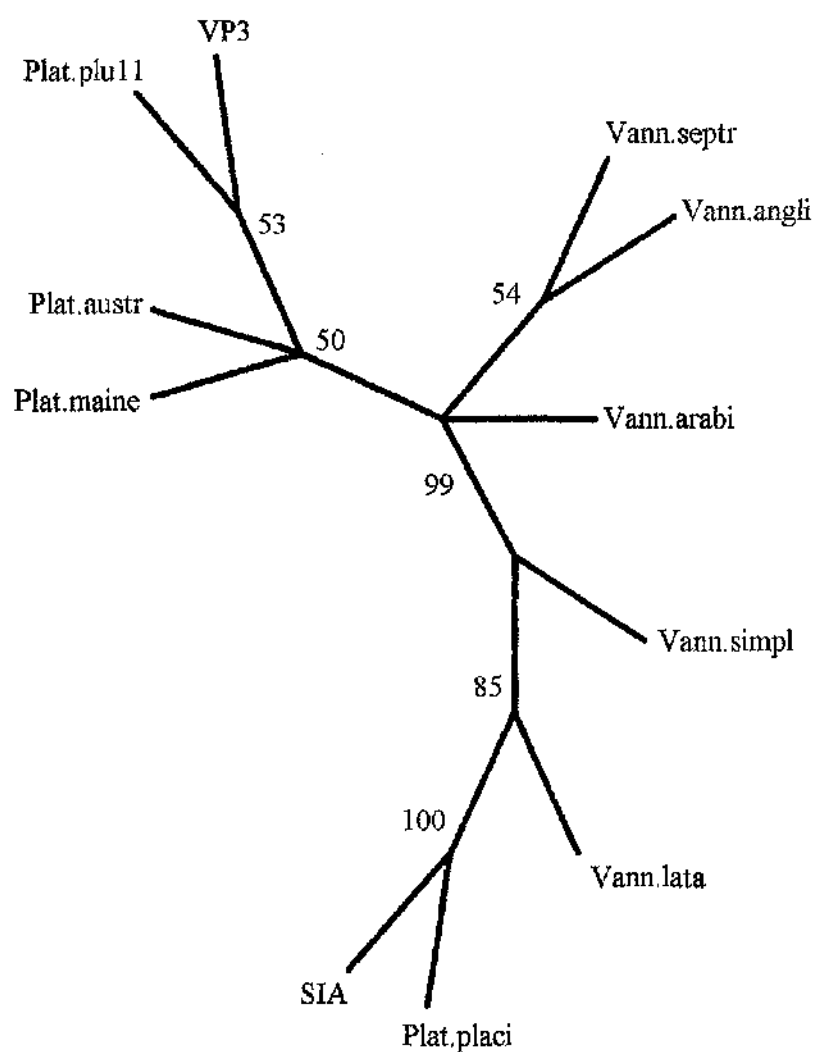
Considering that the Species E produced clearly the most unusual sequence it was excluded from the data set and a bootstrapped parsimony analysis was performed with the remaining taxa. The unrooted, majority-consensus most parsimonious tree is shown in Figure 5.7. The branching order of the isolates was not affected by the removal of Species E from the analysis, but the bootstrap values did alter considerably. The bootstrap support for the separation of the marine and freshwater clusters was 99%, and there was also weak support for the separation of *Vannella* and *Platyamoeba* marine isolates (50%), the association of *V. septrentionalis* and *V. anglica* (54%), and the *P. plurinucleolus* and VP3 (53%). This would suggest that although the presence of Species E may reduce the confidence of some of the branches it does not otherwise affect the topology of the tree.

### Rooting the tree

All the phylogenetic trees which have been presented up to this point are unrooted so there is no indication of which taxa are the most ancient and which are the most recent. Analysis of the sequence data demonstrates that Species E has the most



**Figure 5.6** Unrooted phylograms of naked, fan-shaped amoebae using a restricted data set from the V7 V8 variable region. (A) Majority-consensus most parsimonious phylogram. The data set was resampled 100 times with the bootstrap method and the most parsimonious trees were determined using a branch and bound algorithm. Bootstrap values less than 50%, and branches occurring in less than 50% of the trees are not shown. The horizontal scale represents one observed nucleotide change. (B) Maximum likelihood phylogram. The horizontal scale represents 10 expected evolutionary changes every 100 nucleotide positions.



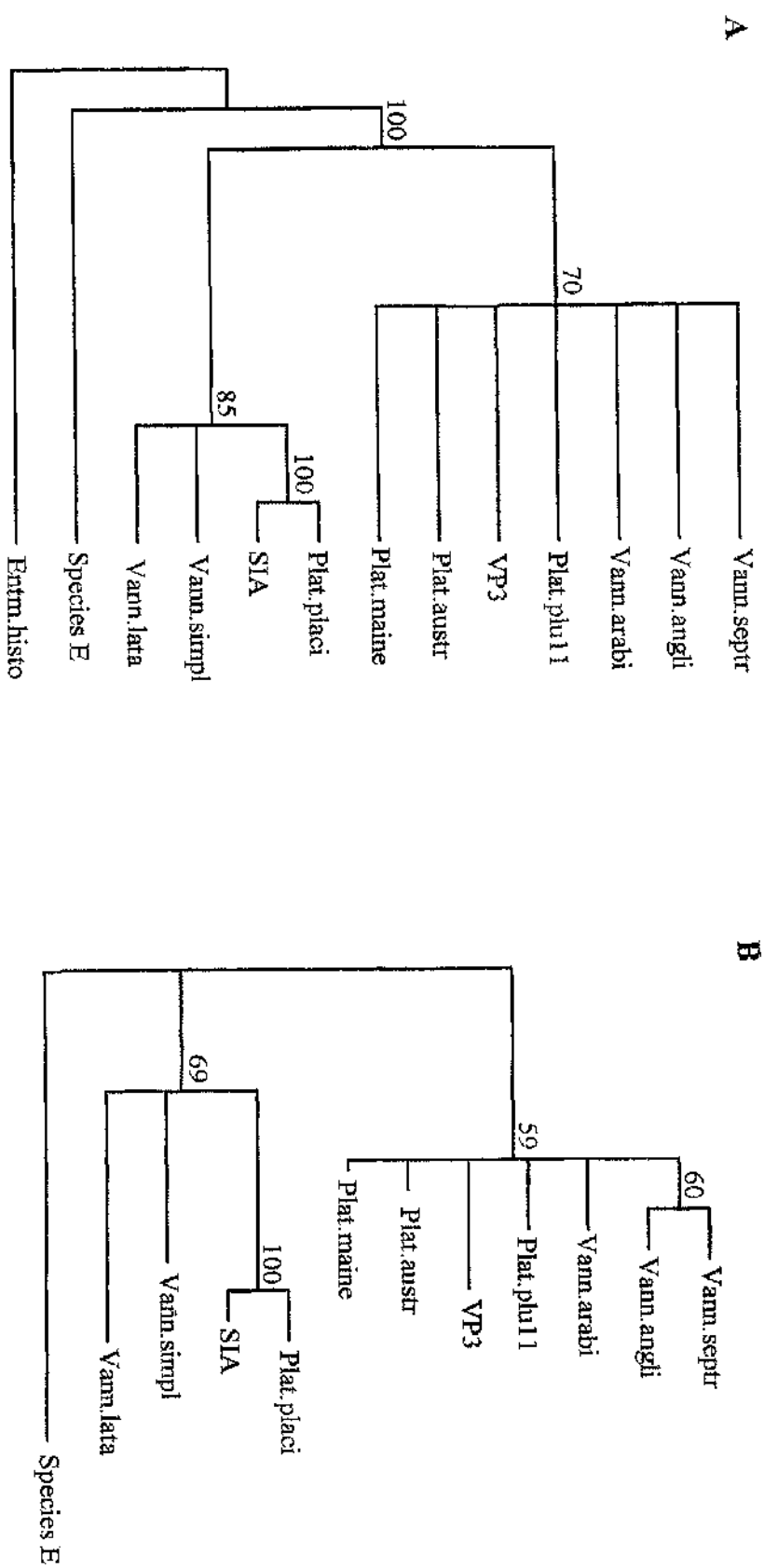
**Figure 5.7** Unrooted, majority-consensus most parsimonious phylogram with the exclusion of the Species E taxon. The data set was resampled 100 times with the bootstrap method. Bootstrap values less than 50%, and branches occurring in less than 50% of the trees are not shown.



unusual sequence and the root of the tree is therefore likely to fall somewhere along its considerably long branch. However, it is also possible that the root may occur elsewhere on the tree possibly between the freshwater cluster and the branch of Species E or between the marine cluster and the Species E branch. To determine the position of the root of the tree, an outgroup taxon, *Entamoeba histolytica*, was introduced into a parsimony analysis. The rooted semi-consensus most parsimonious tree is shown in Figure 5.8 A. Rooting the tree with the *E. histolytica* resulted in bootstrap support of 70% for the freshwater cluster and 85% for the marine cluster. The 100% support for the association of Species E and *E. histolytica* demonstrates that Species E is an outgroup with respect to all the *Vannella*, *Platyamoeba* and unidentified isolates, and the root of the tree lies along the long Species E branch. When the parsimony and maximum likelihood trees were rooted with Species E as an outgroup, the branching of the *Vannella* and *Platyamoeba* isolates only altered with respect to Species E, not with respect to each other. Consequently, the bootstrap values for the rooted majority-consensus most parsimonious tree, shown in Figure 5.8 B, differed, although only marginally, from the unrooted tree previously presented in Figure 5.6 A. For the rooted phylogeny, there was 59% and 69% bootstrap support for the marine and freshwater clusters, respectively. Within the freshwater cluster there was no bootstrap support for the branching order of *V. lata* and *V. simplex*, but there was 100% support for the association of *P. placida* and the unidentified SIA isolate. Within the marine cluster, there was 60% support for the association between *V. anglica* and *V. septrentionalis*, although the branching order among the remaining marine isolates was unresolved.

### Assessment of the usefulness of diagnostic features

A series of constraint trees were constructed which split the isolates into two or three groups on the basis of a morphological or ultrastructural feature. Using these constraints, the usefulness of these features as phylogenetic markers was assessed by comparing the lengths of the most parsimonious trees. Using the cell surface structure as a diagnostic feature produced a most parsimonious tree with a total length of 198. Size of the locomotive form, represented by the average size of the largest dimension, produced a tree length of 202. Shape of the locomotive form, depicted, by the length:breadth ratio produced a tree of 200 evolutionary changes. Finally, the length of the floating form pseudopodia, as a proportion of the size of the central mass, produced a tree length of 204 evolutionary changes. These tree lengths are all relative to the unconstrained most parsimonious tree length of 183,



**Figure 5.8** Rooted, majority-consensus most parsimonious phylograms. (A) Rooted with *Entamoeba histolytica* as an outgroup. (B) Rooted with Species E as an outgroup. The data set was resampled 100 times with the bootstrap method and the most parsimonious trees were determined using a branch and bound algorithm. Bootstrap values less than 50%, and branches occurring in less than 50% of the trees are not shown.

which rather unexpectedly, divided the isolates with respect to the habitat from which they were isolated. The most parsimonious tree, which divided the isolates with respect to habitat is clearly significantly better than when the isolates are divided with respect to the cell surface structure, the character which is currently used to classify these morphologically similar isolates into the genera *Vannella* and *Platyamoeba*. Moreover, this result suggests, that cell surface structure is only marginally more useful as a phylogenetic marker than shape or size of the locomotive form, or the size of the floating form pseudopodia.

### Identification of unidentified isolates

Previous light microscopical observations on the locomotive form, floating form and cysts of isolate SIA were very similar to *P. placida* (Chapter 2). Throughout the phylogenetic analyses using partial *ssrRNA* sequence data, there was also strong support for a relationship between *Platyamoeba placida* and SIA. Together these results would suggest that SIA is closely related to *P. placida*, and most likely another strain of this species.

The identity of the other two unknown isolates, Species E and VP3, is far less certain. Light microscopical observations, riboprinting and partial *ssrRNA* sequence analyses all suggest that VP3 is a marine *Vannella* or *Platyamoeba* isolate. Light microscopical observations suggested that VP3 was most similar to *V. arabica*. Generally riboprinting data and partial *ssrRNA* sequence analyses were unable to resolve a clear relationship between VP3 and any other marine isolate, although there were weak associations with *V. anglica* and *P. australis* in the riboprinting analysis, and *P. plurinucleolus* in the sequence analysis.

With respect to size, Species E resembled some small *Platyamoeba* isolates and *V. aberdonica*. From observations of the locomotive form, Species E was considered to be a strain of *V. aberdonica* (Butler, 1994), however, the length: breadth ratio of Species E is much greater than any of these isolates, as a consequence of its tendency to adopt a linguiform shape. Moreover, the pseudopodia of the floating form of Species E were different from any of the described amoebae of this size. Although molecular data is available for *V. aberdonica* or any of the small *Platyamoeba* species, riboprinting results suggested that Species E may be quite different from other *Vannella* and *Platyamoeba* isolates, and the partial *ssrRNA* sequence analysis demonstrated that Species E was clearly an outgroup to these

other morphologically similar isolates.

### **Examination of ultrastructural features**

To further investigate the identities of VP3 and Species E, ultrastructural features of these isolates were examined by transmission electron microscopy. Specifically, the nucleus was examined to establish whether the nucleolar material was central or arranged in parietal lobes, the mitochondria were examined to determine the shape of the mitochondrial cristae, and the cell surface structure was examined for the presence of an undifferentiated glycocalyx or distinct glycostyles.

### **VP3**

Transmission electron microscopy clearly showed that VP3 possess a nucleus which ranges from 2.5-4  $\mu\text{m}$  in diameter with a central nucleolus of approximately 0.8  $\mu\text{m}$ . The mitochondria were spherical in shape with branched tubular cristae (Figure 5.9 A). The cell surface was composed of a dense basal layer of 10-12 nm in thickness and filamentous elements which extended approximately 40 nm from the cell membrane (Figure 5.9 B and C). The 'glycostyles' were not numerous or obvious, and they were only apparent when the membrane was sectioned at a favourable angle. The glycostyles did have the characteristic longitudinal shape, although the shape of the cross-section was not discernable.

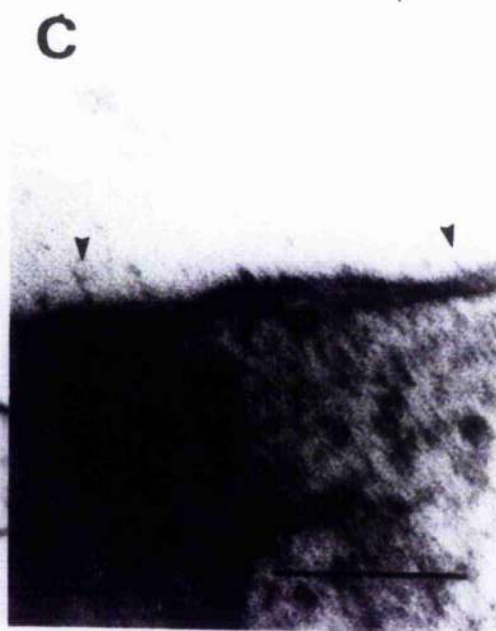
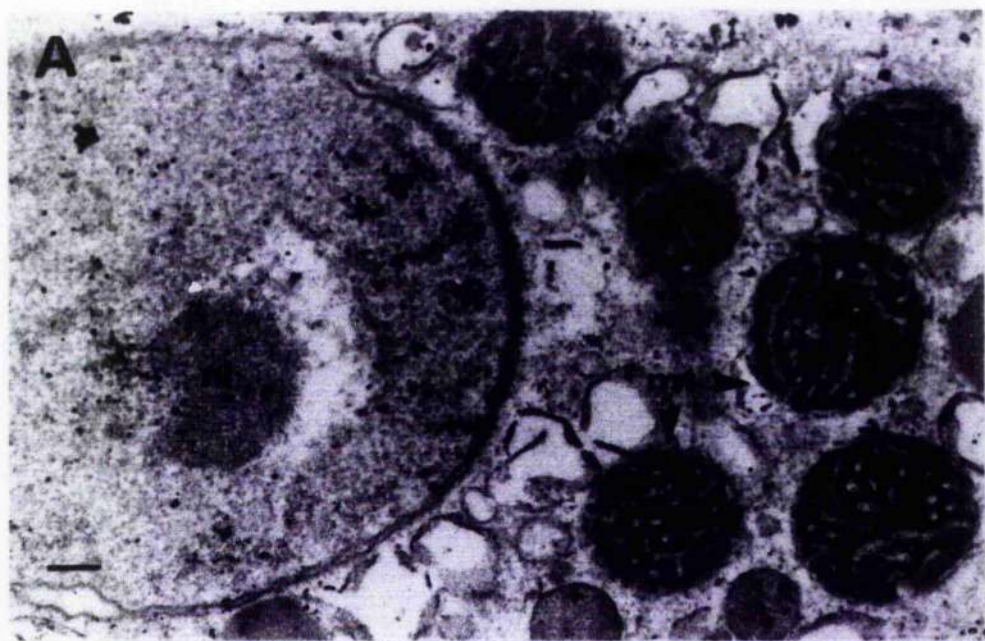
---

**Figure 5.9** Electron micrographs of unidentified isolate VP3.

(A) The nucleus with a central nucleolus (nl), and spherical mitochondria (mt) with branched tubular cristae.

(B and C) The fine structure of the cell membrane with glycostyles (arrowed).

The scale bars = 0.2  $\mu\text{m}$  in each case.



This is contrary to some reports of *Vannella* spp. which have shown the glycostyles to be obvious features at relatively low magnification and clear in cross-section showing their pentagonal arrangement (Page, 1983). As a control to test the effects of the TEM preparation protocol, the cell surface structure of *V. anglica* was examined (Figure 5.10 A and B). The surface structure was composed of a compact basal layer of 7.5 nm which was separated from the plasma membrane by 5 nm, with filamentous elements of approximately 20 nm. The presence of numerous 'glycostyles' was more obvious with this isolate, however, individual 'glycostyles' were smaller than other described *Vannella* species. Despite being small, the characteristic longitudinal shape of some 'glycostyles' was discernable, which would suggest that they have not been truncated during preparation. No simple filaments among the glycostyles were detected. A previous published examination of the cell surface structure of *V. anglica* showed that the glycostyles were up to 126 nm in length with simple filaments to approximately 240 nm (Page, 1980a). This suggests that either the large glycostyles are removed during different fixation or staining procedures, or that the length of glycostyles may vary with the availability of food, the time spent in culture or the culture media. The same fixation and staining procedure was used to examine glycostyles and filaments of another undescribed *Vannella* isolate which are up to 500 nm in length (Figure 5.10 C). This supports the view that the TEM procedures used in this study were appropriate for observing glycostyles, and that strains may alter their cell coat features when kept in culture for long periods.

---

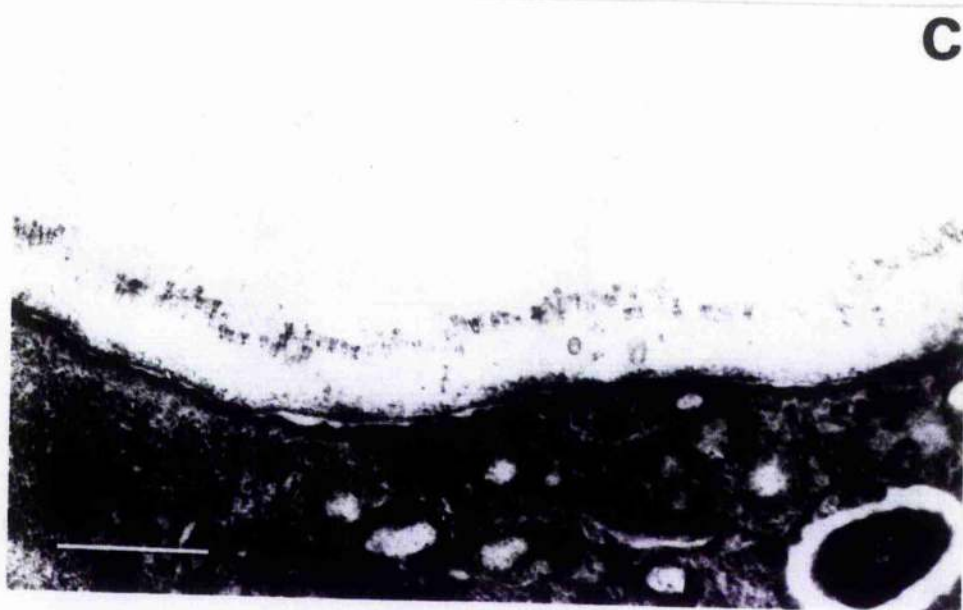
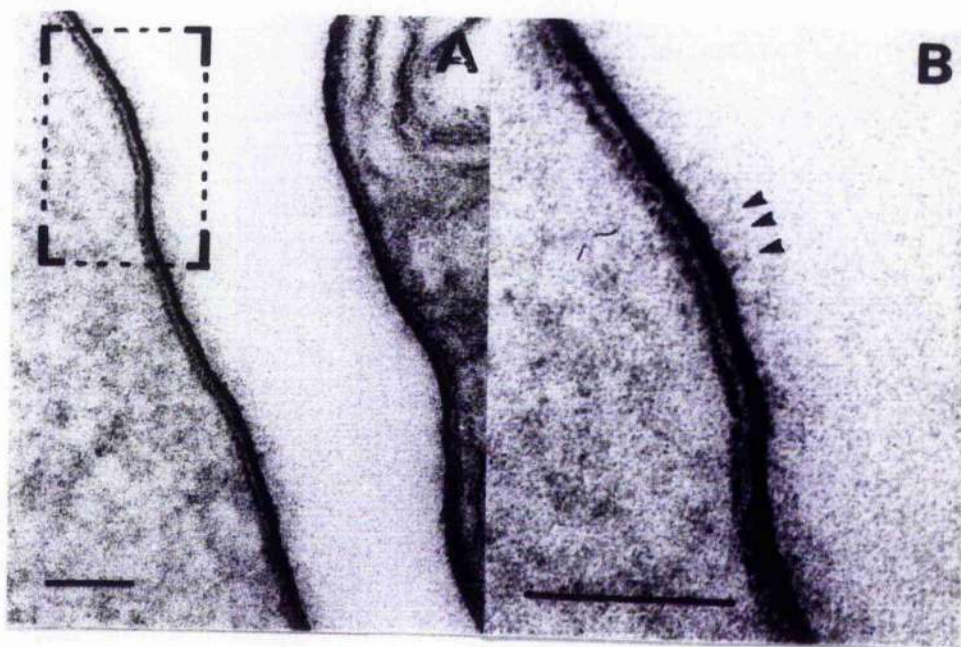
**Figure 5.10** Electron micrographs of the fine structure of the cell membrane from *Vannella* species.

(A) The cell membrane structure of *Vannella anglica* showing small glycostyles of approximately 20 nm. Scale bar = 0.1  $\mu$ m.

(B) An enlargement of with the most obvious glycostyles indicated with arrows. Scale bar = 0.1  $\mu$ m.

(C). The cell membrane of an unknown *Vannella* sp. showing large glycostyles of approximately 500 nm in length (courtesy of Evelyn Armstrong). Scale bar = 1  $\mu$ m.





## Species E

Examination of sections at low magnification showed that a large proportion of the cell was taken up by food vacuoles. The shape of the nucleus was variable with a mean diameter of 2  $\mu\text{m}$ . The mitochondria were oval in shape with an average largest dimension of 720 nm (Figure 5.11 A). Examination of Species E at higher magnifications clearly demonstrated that the nucleus contained a central nucleolus of approximately 325 nm (Figure 5.11 B). No endoplasmic reticulum was associated with the nucleus. Closer examination of the mitochondria showed that they possessed branched tubular cristae (Figure 5.11 C). Small, spherical densely stained microbodies of unknown origin, approximately 175 nm in size were also occasionally encountered. Despite examination of over a hundred cells there was no evidence to suggest that Golgi apparatus were differentiated into distinct dictyosomes. Examination of the cell surface structure demonstrates that the glycocalyx is very thin. It is composed of a compact smooth layer of 1-2 nm, with filamentous material completely absent, separated from the plasma membrane by 1-2 nm.

---

**Figure 5.11** Electron micrographs of unidentified isolate Species E.

(A) Whole cell showing the nucleus (n), mitochondria (mt) and food vacuoles (v). Scale bar = 1  $\mu\text{m}$ .

(B) Cell nucleus with central nucleolus (nl). Scale bar = 0.2  $\mu\text{m}$ .

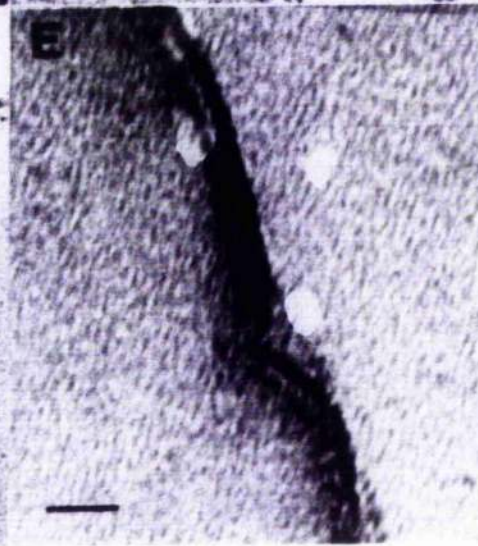
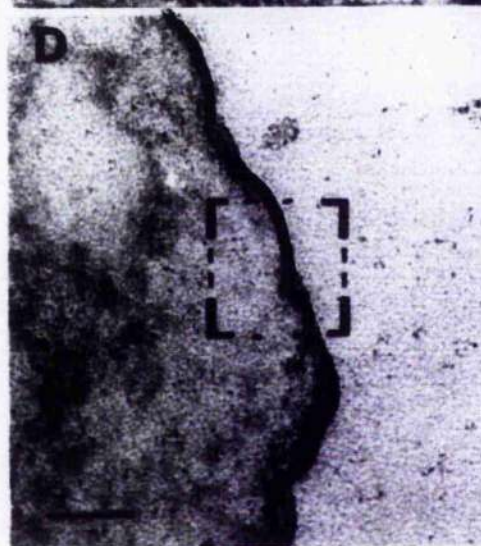
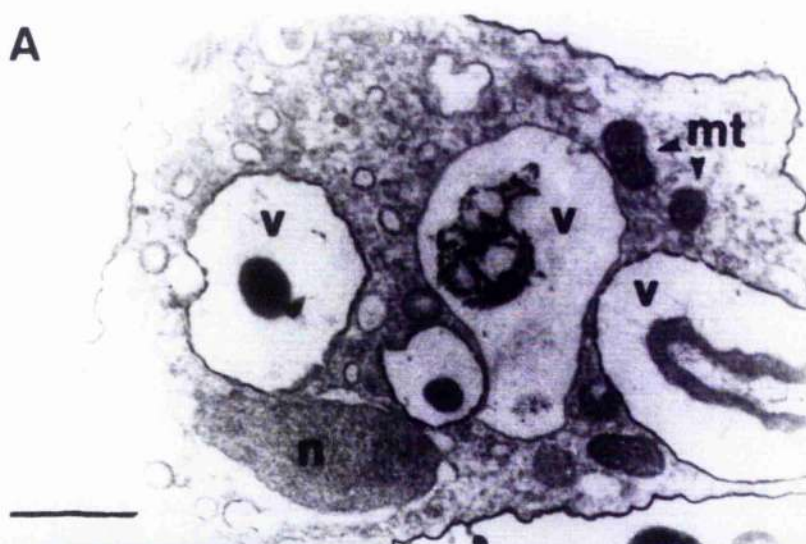
(C) Mitochondrion with branched tubular cristae, and a densely stained microbody (mb). Scale bar = 0.2  $\mu\text{m}$ .

(D) Fine cell surface structure without glycostyles photographed at x 40,000. Scale bar = 0.1  $\mu\text{m}$ .

(E) Enlarged region from (D) photographed at x 100, 000. Scale bar = 10 nm.



A



## Discussion

### **Length of V7 V8 ssrRNA region**

Sequence analysis demonstrated that the size of the V7 V8 ssrRNA gene fragment varied from 380 bp in *Platyamoeba placida* to 438 bp in Species E. This size is characteristic of the V7 V8 region from other eukaryotic organisms. A comparison of 17 sequences from a variety of eukaryotic taxa demonstrated that the size of this region ranges from 275 bp in the microsporidian *Encephalitozoon cuniculi* (Pieniazek, 1993b), to 590 bp in the anaerobic naked amoeba *Phreatamoeba balamuthi* (Hinkle *et al.*, 1994). However, a majority of taxa, including *Saccharomyces cerevisiae*, *Homo sapiens*, *Babesia bigemina* and *Oxytrichia granulifera*, were of a similar size to the fan-shaped amoebae, with a range in size from 367-403 bp (Rubstov *et al.*, 1980; McCallum & Maden, 1985; Reddy *et al.*, 1991; Schlegel *et al.*, 1991).

The comparatively small V7 V8 ssrRNA region of *E. cuniculi* can be attributed to the absence of conserved elements which are almost universal among the eukaryotes (Hartskeerl, *et al.*, 1993). Helix 43 of *E. cuniculi* is very short, and helices 44 and 46 are completely absent. The expansion region between helices 45 and 46, which is present in some eukaryotes is also absent. Together almost the entire assemblage of hyper-variable positions are absent from this taxon. The fan-shaped amoebae, like the other eukaryotic sequences analysed, possessed all the conserved helices from 38 to 48 in the V7 V8 region.

The comparatively large V7 V8 ssrRNA regions of *Phreatamoeba balamuthi*, and other taxa such as *Acanthamoeba castellanii* and *Plasmodium falciparum* (Dams, *et al.*, 1988; Neefs, *et al.*, 1990), were generally a result of a large extended helix 43 and an exceptionally large expansion segment between helices 45 and 46. For these three taxa, helix 43 was from 100-130 bases larger than a majority of other eukaryotes, and the expansion segment which varied between 47 and 77 bases and is either absent or small in most other eukaryotes. Moreover, among the eukaryotic taxa, an extended region of approximately 50 bases within helix 44 was also identified, although this was unique to *Entamoeba histolytica* (Que & Reed, 1991).

The majority of the size variation among fan-shaped amoebae was also confined to helix 43 and the expansion segment between helices 45 and 46. Helix 43 of Species E was approximately 60 bases longer than the other fan-shaped amoebae and most other eukaryote sequences. The region between helices 45 and 46 was five nucleotides in size for Species E, which is typical for most eukaryotes, but considerably larger in all the other isolates sequenced ranging from 13 to 23 nucleotides. The expanded region of these isolates was also rich in adenines and thymidines residues.

### **Secondary structure of expansion segment E45-1**

In general, the secondary structure of the AT-rich expansion segment could not be determined, and the sequences could not be unambiguously aligned. For most taxa, the high proportion of thymidines ensured the expansion segment could not form a helix, although a small helix involving five paired residues may be present in *Vannella arabica*, *V. septentrionalis*, *V. lata* and VP3 (Chapter 3). The secondary structure of the AT-rich expansion segment has not been determined for *E. histolytica*, although it forms a single helix with a large bulge in *P. falciparum*. A large single helix is also apparent in the expansion segments of *A. castellanii* and *P. balamuthi*, which have a more balanced base composition (Dams *et al.*, 1988; Hinkle *et al.*, 1994).

### **Sequence variation**

Examination of the sequence variation among the fan-shaped amoebae demonstrates that Species E is considerably different from any other isolate. Sequence comparisons among the remaining isolates showed that the variation among the marine isolates is much lower than the sequence variation among the freshwater isolates. Generally, the sequence variation detected among the marine isolates in this study is much greater than the interspecific variation for other organisms. An examination of nine species from the yeast genus *Zygosaccharomyces* identified sequence variation in the order of 0.2-2% which is equivalent to 3-35 base differences over the complete *ssrRNA* gene (James *et al.*, 1994). In the case of some *Tetrahymena* species, no sequence differences were determined at all (Sogin *et al.*, 1986c).

In some organisms the degree of sequence variation is high. Partial *ssrRNA* gene sequences from the V7 region from species of the trypanosome genus *Phytomonas*

identified 55 parsimony informative positions (Marché *et al.*, 1995), which is compared to 56 for the restricted V7 V8 data set used in the analyses presented here. Partial sequence analysis of the V7 V8 region from *Acanthamoeba* species demonstrated that the sequence diversity was marginally greater than the variation detected among the marine isolates, approximately equivalent to the variation between vertebrates and invertebrates (Johnson *et al.*, 1990). However, the sequence diversity of the freshwater isolates was considerably greater than the marine isolates and almost all other examples of interspecific variation. This scale of sequence variation has been exhibited among the amoeboid-flagellate genus *Naegleria*. Within a consecutive domain of 801 nucleotides, which encompassed a majority of the V2, V3 and V4 regions, there were three differences between clusters of the *N. gruberi* species complex and 96 differences between *N. andersoni* and *N. jadini* (De Jonckheere, 1994b).

Despite the sequence variation among the *Vannella* and *Platyamoeba* species, one sequence element in the V7 region was identical for all *Vannella* and *Platyamoeba* species. Comparisons with Species E and other taxa identified three or more differences over this region. Therefore, this site may be useful for the direct identification of *Vannella* and *Platyamoeba* species in mixed cultures using *in situ* hybridization, or indirect identification using hybridization with extracted nucleic acids.

### Phylogenetic analyses

Phylogenetic analysis of the fan-shaped amoebae demonstrated that the freshwater isolates and the marine isolates formed two clades separated by a diverse unidentified marine isolate Species E. Within the freshwater cluster, the *Vannella* isolates branched before the *Platyamoeba* and unknown isolate SIA. These findings were consistent, irrespective of the phylogenetic method employed and whether the complete data set was used or a more restricted data set, with ambiguously aligned sites removed. The branching order among the marine isolates could not be resolved.

For the unrooted phylogenies, the bootstrap support for the parsimony phylogeny was generally quite weak ranging from 58-65% for the freshwater clade and remaining stable at 71% for the marine clade. Some phylogenies can be affected by individual taxa, but removal of Species E which was suspected of having a significant influence on the tree, did not alter the topology and the bootstrap

support for the separation of the freshwater and marine isolates actually increased to 99% in its absence. The introduction of outgroups can upset weak phylogenies, but this did not occur when *E. histolytica* or Species E were used to root the tree. For these rooted phylogenies, the bootstrap support for the freshwater and marine clades ranged from 69-85% and 59-70%, respectively. So despite consistent separation of the freshwater and marine clades in all these analyses, the bootstrap values were generally low, and the phylogenetic division of these clades is not completely convincing. The inclusion of more taxa in the analysis, and perhaps, additional sequence data from other variable regions of the ssrRNA is required to confirm these results.

In support of this molecular phylogeny, there does appear to be a correlation between the genetic variation and the morphological variation exhibited by the freshwater and marine isolates. The size and shape of the locomotive form of the marine isolates is generally similar, and although the length of the floating form pseudopodia varies, they are all rather irregular and blunt. Conversely, the locomotive form of the freshwater isolates is more variable with respect to both size and shape, and morphology of the floating form pseudopodia, and the presence or absence of cysts.

Because of the low genetic variation between the marine isolates, it was not possible to determine the phylogenetic relationships between these isolates. In addition, there was no noteworthy morphological variation among the marine isolates suggesting that these isolates, with the exception of Species E, are a monophyletic group. Considering the morphological similarities between *Clydonella* species, and marine *Vannella* and *Platyamoeba*, it is probable that if *Clydonella* species were included in this analysis they would also fall within this genetically monophyletic group of marine amoebae. This would support the rejection of the *Clydonella* genus in the absence of any ultrastructural analysis (Page, 1987). Phylogenetic analysis using sequence data from the remaining ssrRNA variable regions may allow relationships among these marine isolates to be resolved.

The variability among the freshwater isolates far exceeds the variation among the marine isolates, and it is less likely that this group is monophyletic. Additional data from freshwater isolates would provide a better indication of the relationships among these isolates. Preliminary sequence analysis from *V. platypodia* suggests

that this freshwater isolate is also considerably different from the marine isolates. The sequence variation among the freshwater isolates suggests that the marine isolates may have evolved from a common ancestor more recently than the freshwater isolates, or that these groups are affected by unequal rates of evolution, with the freshwater isolates evolving more quickly than the marine isolates. If the latter is true, the monophyly of the marine isolates may be an artifact caused by branch length attraction of the freshwater isolates (Felsenstein, 1978). If there was an unequal rate of evolution between the freshwater and marine isolates, the base substitutions between these groups should be random and independent: examination of the alignment shown in Figure 5.4 demonstrates that this was not the case. At eight positions, numbered 86, 97, 348, 394, 502, 503, 511 and 512, three of the four freshwater isolates have the same base substitution from all the marine isolates, and at two positions numbered 251 and 346 all four freshwater isolates have the same nucleotide substitution from the marine isolates. Together these substitutions, which account for one fifth of the total number of parsimony informative sites, support the notion that the marine isolates diverged recently, and the freshwater isolates have a more ancient origin.

The phylogenetic division of the 12 fan-shaped amoebae with respect to habitat was a rather unexpected result, as these organisms have been classified, on the basis of the cell surface structure, into two genera, *Vannella* and *Platyamoeba*. Freshwater isolates and marine isolates can be distinguished using two morphological characters. Freshwater isolates produce cysts or have long floating form pseudopodia which taper to a point. No cysts have been described for any marine isolate, and although some isolates have long floating form pseudopodia, none taper to a point. Small-subunit ribosomal sequence comparisons with isolates of other genera may identify close phylogenetic relationships with other amoebae with long, pointed pseudopodia or smooth cysts.

Within the marine clade, all the isolates were isolated from coastal or estuarine waters. Species E was the only marine isolate not to fall within the marine clade. Interestingly, this isolate was isolated from marine sediment at a water depth of 43 m, perhaps a quite separate habitat within the marine environment. A survey on the diversity of naked amoebae within marine sediments demonstrated that this habitat is host to several small flat fan-shaped isolates which could not be allocated to a described species (Butler, 1994). Although, these benthic amoebae are



comparatively small, there are otherwise no obvious morphological features which differentiate them from the other marine isolates. The general lack of distinguishing morphological features of marine fan-shaped amoebae has been an inherent problem for identification and the phylogenetic examination, but these results suggest that there may be considerable genetic differences between small fan-shaped benthic amoebae and previously described *Vannella* and *Platyamoeba* species. The marine benthos may host a whole assemblage of undescribed amoebae.

The phylogenetic segregation of the isolates cannot be attributed to the geographical location from which they were isolated. Although a majority of the marine isolates including Species E were isolated from Britain, *V. arabica* was isolated from Kuwait, *P. mainensis* was isolated from North America and *P. australis* from Australia. Among the freshwater isolates, *V. simplex* and *V. lata* were isolated from Germany and Scotland, respectively, whereas the very similar *P. placida* and SIA were from North America and England. Characteristic flat, fan-shaped amoebae have also been isolated from coastal waters and freshwater lakes in Antarctica (Butler, unpublished; Mayes, 1995). There is therefore no correlation between geographical distribution and the molecular phylogeny determined from partial *ssrRNA* sequence data.

Most parsimonious trees were determined for a series of constraint trees which divided the isolates into two or three groups with respect to a morphological or ultrastructural feature. When the cell surface structure was used as a diagnostic feature the most parsimonious constraint tree had a length of 198 evolutionary changes, considerably longer than the most parsimonious unconstrained tree which had a length of 183. This would suggest that the cell surface structure is less useful than the habitat or perhaps the combination of the presence of cysts or floating form pseudopodia which taper to the point, as a phylogenetic marker. Moreover, the use of shape, quantified by the length: breadth ratio, greatest dimension and the length of the floating form pseudopodia were not substantially worse as phylogenetic indicators than the cell surface structure. The greatest dimension and the shape of a population of cells may vary significantly with culture conditions, and the length categories for the floating form pseudopodia can be considered very subjective. These features were therefore unlikely to be accurate phylogenetic markers, but the constrained trees were only two, four and six evolutionary changes longer than the most parsimonious tree constrained by cell surface structure, which is considered

a rigid diagnostic character. The usefulness of other diagnostic features such as the nature of the nucleolar material and the mitotic behaviour were not examined.

In all *Vannella* and *Platyamoeba* species described, the nucleolar material is in the form of a central nucleolus, except *V. devonica* and *P. plurinucleolus* which have nucleolar material arranged in parietal lobes. Of these two species *P. plurinucleolus* was sequenced, and associated with all the other marine isolates which suggests this character is not an important phylogenetic marker, but is useful for identification of these particular species. Whether the presence of a parietal nucleolus evolved independently for *V. devonica* and *P. plurinucleolus* among these morphologically similar fan-shaped amoebae is uncertain.

There is considerable variation in the mitotic behaviour among species of *Vannella* and *Platyamoeba*. In all cases, there are no centrioles or other microtubular organising centres present and the nucleolus disappears during prophase, although there is considerable variation in the degree, and stage, that the nuclear membrane disintegrates (Page, 1983). The nuclear membrane has usually disintegrated by metaphase although it may persist much longer and remnants of it may form polar caps. The usefulness of this diagnostic feature as a phylogenetic marker was not examined as details of the mitotic process have not been described for many species and the fate of the nuclear membrane at any particular stage is not certain. Previous examinations would suggest that this character is not useful for distinguishing between marine isolates of *Vannella* and *Platyamoeba*, however, it is more useful for distinguishing between the freshwater species (Page, 1980a).

### **Re-evaluation of cell surface structure**

Although there are sufficient morphological differences between freshwater species to assign fan-shaped amoebae isolated from the field to a particular species of *Vannella* or *Platyamoeba*, the morphological differences between marine species are far less obvious. Flat fan-shaped amoebae without trailing uroids and subpseudopodia are classified into the genera *Vannella* and *Platyamoeba* on the basis of the fine surface structure. *Vannella* are described as possessing a glycocalyx differentiated into distinct glycostyles and *Platyamoeba* are described as having a non-distinct 'fuzzy' glycocalyx. However, clear differences in cell surface structure between these morphologically similar genera are not always apparent. Therefore, it may be time to review the use of surface structure as the principal character in



identification of these genera.

Analysis of the cell surface structure of *V. anglica* and VP3 in this study, demonstrated these isolates had distinct filamentous elements of 20 and 40 nm respectively. These elements were significantly smaller than glycostyles from other *Vannella* species which are usually between 100-130 nm in size (Page & Blakey, 1979; Page, 1979a, 1980b). The size range of glycostyles previously reported may have been underestimated as recently these features have been shown to vary from 45 nm (Butler, 1994) to 500 nm (Armstrong, unpublished data) in morphologically similar undescribed isolates. However, a previous examination of the cell surface of *V. anglica* showed that glycostyles extended from 117-126 nm from the plasma membrane with simple filaments to about 240 nm (Page, 1980a). In this study, the filamentous elements of both *V. anglica* and VP3 were of a comparable shape to glycostyles and they did not appear to be truncated. This suggests that for *V. anglica* at least, the size of the glycostyles may vary, perhaps as a result of long periods in culture, or the filamentous elements examined were not glycostyles but discrete elements which comprise the glycocalyx.

The glycocalyx of *Vannella* isolates is usually described as simply amorphous (Page & Blakey, 1979), although its compact form with less dense filamentous regions bear the characteristic features of *Platyamoeba* species. Of the *Platyamoeba* species previously examined by electron microscopy, the compact layer of the glycocalyx varied in thickness from 5-10 nm and the range of thickness of filamentous layer was 10-40 nm (Page & Blakey, 1979; Page 1980b), although *Platyamoeba*-like isolates have glycocalyces as thick as 60 nm (Butler, 1994). Closer examination of the fibrous region of *Platyamoeba* glycocalyces suggest they are arranged in rows of elements which are usually tightly packed and visually inseparable, although there is a suggestion of a central filamentous core (Page, 1980b). Moreover, the fibrous region has been shown to be variable between strains of the same species, and affected by different fixation methods (Page & Blakey, 1979). It has also been noted that glycostyles may be removed or destroyed from some strains during fixation by the use of long incubations with osmium (Page, 1983). Together these results suggest that the fixation step of TEM preparation is critical in the examination of the cell surface structure of these amoebae. The fixation procedure may remove or destroy glycostyles, and affect the appearance of the glycocalyx even within strains of the same species. The distinct differences in the fine cell surface

structure between *Vannella* and *Platyamoeba*, may therefore, be attributable to TEM artifacts. A similar case may exist in another gymnamoebae genera, *Vexillifera*. Marine isolates within this genus possess discrete glycostyles in the form of hexagonal cylinders which extend from the plasma membrane from 50-70 nm (Page 1979b, 1983). Both the freshwater isolates so far examined do not have distinct glycostyles, but have an indistinct layer of cylindrical elements of 10 and 17 nm in size (Page, 1988). In perhaps both these cases, the cell surface structure should not be regarded as a rigid diagnostic character, rather a character which may be useful for identification.

### Identification of SIA

Examination of the morphological characters of SIA immediately demonstrated that of all the described freshwater *Vannella* and *Platyamoeba* isolates, SIA only resembled *P. placida*. Indeed, there were no significant differences between these two isolates despite being isolated from different sides of the Atlantic Ocean. Comparisons of the partial ssrRNA sequences demonstrated that there were six differences between SIA and *P. placida* for the restricted data set, fewer than between any other pair of isolates. Furthermore, the phylogenetic analyses associated *P. placida* and SIA in all parsimony analyses with 100% bootstrap support, and maximum likelihood analyses. Together, these results provide strong evidence that the freshwater isolate SIA is a strain of the *Platyamoeba* type species, *P. placida*. Confirmation using transmission electron microscopy was regarded as unnecessary.

### Identification of VP3

Phylogenetic analysis of the partial ssrRNA sequences demonstrated that VP3 was a typical marine *Vannella* or *Platyamoeba* isolate. Although no clear relationship was apparent with any particular isolate, VP3 most frequently associated with *P. plurinucleolus*. Comparisons of morphological features demonstrated that VP3 superficially resembled a number of *Vannella* isolates and *P. bursella*, although it was most similar to *V. arabica*. Ultrastructural examination using transmission electron microscopy demonstrated that VP3 has branched tubular mitochondrial cristae and a central nucleolus. Mitochondrial cristae of this type are common to all *Vannella* and *Platyamoeba*, and a majority of isolates have a central nucleolus. Two isolates, *V. devonica* and *P. plurinucleolus* do not have central nucleoli, instead their nucleolar material is arranged in lobes around the periphery of the nucleus (Page

& Blakey, 1979; Page, 1979a). This clearly demonstrates that VP3 is not an isolate of either of these species.

Examination of the cell surface structure of VP3 has identified a compact layer of 10-12 nm and filamentous elements or 'glycostyles' which extend approximately 40 nm from the membrane. The 'glycostyles' did not appear to be truncated as they were of a comparable shape to glycostyles from other *Vannella* species. The 'glycostyles' of VP3 are considerably shorter than those reported for other described *Vannella* species (Page, 1979a, 1980a), although glycostyles as small as 45 nm have been reported for other undescribed *Vannella*-*Platyamoeba*-like isolates (Butler, 1994). As described previously, these 'glycostyles' may be discrete filamentous elements of the glycocalyx, and if glycostyles with a size characteristic of *Vannella* species were present they may have been removed during the fixation process. Due to this uncertainty, VP3 cannot be assigned to either the *Vannella* or the *Platyamoeba* genus on the basis of cell surface structure. The observations of the locomotive form, described in Chapter 2, would suggest that it is most likely an undescribed *Vannella* isolate.

### Identification of Species E

Comparisons of the sequence differences and the phylogenetic analyses of the ssrRNA V7 V8 partial sequencing data demonstrated that Species E was considerably different from other *Vannella*, *Platyamoeba* and unknown isolates. Comparisons of morphological features of the locomotive form examined by light microscopy suggested that Species E was superficially similar to *V. aberdonica* (Page, 1980a), and some small *Platyamoeba* isolates isolated from the surface waters of Chincoteague Bay, Virginia (Sawyer, 1975b). However, the locomotive form of Species E was often more linguiform than these isolates, and light microscopical observations of the floating form also suggested Species E was significantly different from other isolates. These results therefore suggest that Species E has not been previously described.

Examination of the ultrastructural details of Species E identified tubular mitochondrial cristae and a central nucleolus which is characteristic of most *Vannella* and *Platyamoeba* isolates, although no Golgi dictyosomes were detected. The absence of Golgi dictyosomes would suggest that Species E is unrelated to *Vannella*, *Platyamoeba* or any other gymnamoebae and that this amoeba represents

an ancient eukaryotic lineage according to a recent classification scheme (Cavalier-Smith, 1993). However, branched tubular mitochondrial cristae have not been found in any ancient taxa, which would suggest either Species E is unique in this respect, or Golgi dictyosomes were over-looked in this isolate. Golgi dictyosomes have been difficult to detect in other taxa and have not been confirmed in several gymnamoebae genera (Gicquaud, 1979; Page, 1983, 1988).

Examination of the cell surface structure demonstrated that Species E was quite unlike any *Vannella* and *Platyamoeba* isolates, which have been examined by transmission electron microscopy. The surface of the cell membrane was completely devoid of glycostyles or a glycocalyx with fibrous elements. The cell membrane was composed of a thin compact glycocalyx of 1-2 nm with no surface features, separated from the plasma membrane by 1-2 nm. Thin glycocalyces without fibrous differentiated elements have been described for *Flabellula*, *Rhizamoeba* and *Acanthamoeba* from the subclass Gymnamoebia, and for schizopyrenid amoebae from the *Gruberella* and *Heteramoeba* genera.

Despite considerable similarity with the cell surface structure of *Heteramoeba dumnonica* (Page, 1983), Species E is quite distinct from the schizopyrenid genera *Gruberella* and *Heteramoeba*. Species E does not have flagellate stages or cysts, locomotion is non-eruptive and the mitochondrial cristae are branched and tubular, not flattened or discoid. The cell surface structure has been determined from two species of the genus *Rhizamoeba*. The glycocalyces of the marine isolate, *R. saxonica*, and the freshwater isolate, *R. australiensis*, are thin and undifferentiated, but biconvex collosomes just beneath the plasma membrane, which are associated with points of adhesion to the substratum, were detected in *R. saxonica* (Cann, 1984; Page, 1988). Like Species E, *Rhizamoeba* are flattened and fan-shaped when they are less active, although they often have filose holdfasts and are limax during rapid locomotion (Page, 1974b). *Acanthamoeba* have a very thin or undetectable glycocalyx, but unlike Species E, *Acanthamoeba* have slender, flexible tapering subpseudopodia (acanthopodia) and cysts. Moreover, no known marine *Acanthamoeba* have been described, although cysts may survive in sea water (Sawyer & Griffin, 1975).

The remaining genera with a thin glycocalyces are *Flabellula* and *Paraflabellula*. The glycocalyces of species within the genus *Flabellula* are variable. The glycocalyx

of *F. criata* is not discernible, the glycocalyces of *F. calkinsi* and *F. trivnavantica* are amorphous and approximately 8-10 nm above the plasma membrane, whereas the dense glycocalyx of *F. demetica* is only 2.5 nm in thickness and separated from the plasma membrane by 4 nm (Page, 1980c). The gap between the plasma membrane and the very thin smooth glycocalyx of *F. demetica* is very similar to Species E. Species E also has many morphological characteristics in common with members of the *Flabellula* genus. The locomotive form of *Flabellula* are flattened, fan-shaped or spatulate, but commonly irregular and often elongate with a length: breadth ratio larger than one. The locomotive form are usually irregularly rounded, and usually without radiating pseudopodia (Page, 1971, 1980c). However, unlike Species E, *Flabellula* often have eruptive cytoplasmic activity, trailing uroidal filaments, and occasionally clefts in the hyaloplasm, which suggests that Species E is not a member of this genus. Similar morphological differences are apparent between Species E and *Paraflabellula reniformis*, the single marine representative of a closely related genus (Page & Willumsen, 1983).

Morphological examination of Species E suggested that Species E was an undescribed fan-shaped isolate which probably belonged to the genus *Platyamoeba*. Riboprinting provided the first indication that Species E may be quite different from other *Vannella* and *Platyamoeba* isolates (Chapter 2). Comparisons of the partial ssrRNA sequences and phylogenetic analyses, supported the riboprinting result, clearly demonstrating that Species E was considerably different from other supposedly related isolates. Finally, transmission electron microscopy has indicated that the cell surface membrane of Species E is composed a very thin glycocalyx with no fibrous elements or glycostyles unlike any *Vannella* or *Platyamoeba* isolate previously described. Despite similarities in the cell surface structure with other genera, morphological differences suggest that a novel genus should be erected for this isolate. Moreover, I suggest this genera should be regarded as intermediate between the fan-shaped genera *Vannella*, *Platyamoeba* which have differentiated glycocalyces or glycostyles and the fan-shaped genera *Flabellula* and *Paraflabullula* which have thin glycocalyces and locomotive forms often with clefts in the hyaloplasm and trailing uroidal filaments. Furthermore, partial sequence analysis of the V7 V8 variable region has also identified sequence elements which could be used to distinguish species E from *Vannella* and *Platyamoeba* species.

Partial ssrRNA sequence analysis has provided a unique opportunity to examine phylogenetic relationships between morphologically similar amoebae. There appears to be good correlation between the genetic and morphological variation among the *Vannella* and *Platyamoeba* isolates. Freshwater isolates have considerably more morphological and genetic variation than the marine isolates, although the rate of morphological evolution is slow nevertheless. The marine isolates probably form a monophyletic group which appear to have evolved more recently than the freshwater isolates. The molecular phylogeny from the partial ssrRNA sequencing has cast considerable doubt over the usefulness of the cell surface structure as a phylogenetic marker and its use in the classification of fan-shaped amoebae. Partial sequence analysis also confirmed the identity of an unknown freshwater isolate SIA, and indicated that although Species E is superficially morphologically similar to other *Vannella* and *Platyamoeba* isolates, it is genetically quite distinct. Examination by transmission electron microscopy confirms that Species E is unlike *Vannella* or *Platyamoeba* species. Comparisons with other genera suggest some similarities with the genus *Flabellula*, alternatively the erection of a new genus to accommodate Species E should be contemplated. Sequence data from *Flabellula* species, and other isolates similar to Species E which have been described as *Platyamoeba*, such as *P. murchelanoi*, *P. leei*, *P. douvresi* and *P. weinsteini*, would provide large insights into the relationships between amoebae. Moreover, the analysis of additional variable regions, particularly V2, V3 and V4 may resolve the branching order and the phylogenetic relationships between the marine *Vannella* and *Platyamoeba* isolates, and similar isolates described as *Clydonella*.

## **CHAPTER 6**

### **General discussion**

Restriction analysis of PCR-amplified *ssrRNA* genes (riboprinting) demonstrated that there was a significant number of restriction polymorphisms among ten species of naked, fan-shaped amoebae belonging to the *Vannella* and *Platyamoeba* genera. Riboprinting has been successfully used to identify a variety of organisms including naked amoebae (Clark & Diamond, 1991b; Schofield *et al.*, 1991; Clark *et al.*, 1995), but in other instances there are insufficient *ssrRNA* sequence differences to detect polymorphisms between species (Johnston *et al.*, 1993; James *et al.*, 1994). The usefulness of this technique therefore depends on the *ssrRNA* sequence variation exhibited within a taxon to be identified. In this study, there was an adequate degree of sequence variation so that a combination of just two restriction enzymes could be used to distinguish between all the named species. However, the two enzymes used were unable to identify unambiguously fan-shaped amoebae isolated from the field. To facilitate the identification of naked, fan-shaped amoebae the riboprinting database could be extended by increasing the number of restriction enzymes and by including more morphologically similar species. This goal may be accomplished reasonably easily with improvement of PCR efficiency. The secondary structure of the *V. anglica* *ssrRNA* gene indicated that the reverse primer was not complementary to the 3' terminus of the gene. Altering this primer, and optimisation of the PCR conditions, should allow rapid amplification and restriction analysis of other naked, fan-shaped amoebae *ssrRNA* genes. Riboprinting could then be extensively used, in conjunction with light microscopy, for the unambiguous identification of clonal cultures of morphologically similar fan-shaped amoebae. This technique could be applied for the identification of other clonally isolated protists such as morphologically similar *Mayorella* and *Dactylamoeba* species which often require electron microscopy for positive identification.

Riboprinting may also be used to infer phylogenetic relationships if there is sufficient sequence heterogeneity, and the size of the *ssrRNA* gene is consistent (Brown & De Jonckheere, 1994; De Jonckheere, 1994a; Clark *et al.*, 1995). With the development of riboprinting data on naked, fan-shaped amoebae, this technique could also be used to examine the phylogeny of these amoebae. However, to obtain reliable phylogenies based on restriction data, at least ten different enzymes should be used, and individual restriction sites should be mapped to overcome the possible misinterpretation of co-migrating, non-homologous restriction fragments, and the violation of assumptions regarding the independence of characters (Dowling *et al.*, 1996). In this study it was found that a greater quantity of reliable phylogenetic



data could be obtained easily using partial *ssrRNA* sequence analysis. This was primarily because PCR-amplification and TA-cloning of the *ssrRNA* gene fragments was extremely efficient, and phylogenetic inference of sequence data was more informative.

Partial sequence analysis of the V7 and V8 variable regions of the *ssrRNA* was used to examine the phylogeny of the flat, fan-shaped *Vannella* and *Platyamoeba* species, and morphologically similar unidentified field isolates. The phylogeny revealed that the unidentified isolate, Species E, was significantly different from all the *Vannella* and *Platyamoeba* species. Furthermore, the phylogenetic analyses suggested that marine *Vannella* and *Platyamoeba* species were more closely related to each other than they were to any freshwater species, irrespective of the genus. This result was rather unexpected considering that the isolates were classified primarily on the basis of the cell surface structure. However, this result supported evidence from electron microscopy which suggested that the fine cell surface structure may not be a rigid diagnostic feature. The glycostyles are very delicate and may be easily removed depending upon the method of fixation. Moreover, there was considerable variation in appearance of the glycocalyx within a species and the size of the glycostyles may differ with the culture conditions. Considering the increasing use of the fine cell surface structure for diagnostic descriptions and identification, it is clear that intraspecific variation, the stability of the cell surface features and the effect of environmental factors should be fully investigated. It appears that the culture methods and fixation procedures for electron microscopy may have to be optimised for individual isolates.

Interestingly, there was considerably more sequence variation among three freshwater species than that between seven marine isolates. This genetic variation mirrored the morphological variation among the freshwater and marine isolates observed by light microscopy, and there is also some morphological evidence for the division of the freshwater and marine species. Freshwater species can be differentiated from the marine species because freshwater species either have usually long, tapering pseudopodia in the floating form or cysts.

Although the genetic division of the marine and freshwater isolates was consistent, the support for this split was not convincing. To substantiate these findings it is necessary to include more sequence data. Scrutiny of the complete *V. anglica*

ssrRNA indicates that the V2 and V3 variable regions would provide a substantial amount of phylogenetic data; primers for the amplification of V2 V3 region were designed and tested. Furthermore, to determine the validity of another naked, fan-shaped genus, *Clydonella*, data on the cell surface structure and ssrRNA sequences are also required. Comparisons with the other similar genera, *Flabellula* and *Paraflabellula*, should also be considered. These amoebae are also naked, flat and fan-shaped, but differ on account of the presence of trailing uroidal filaments in most cases (Page, 1988).

The use of small-subunit ribosomal RNA sequences has revolutionised our understanding of the evolution of the eukaryotic organisms. Molecular phylogenies using the ssrRNA have generally supported phylogenies based upon rigid morphological characters, confirming the monophyly of a variety of groups including the alveolates, fungi and metazoa (Gajadhar *et al.*, 1991; Van de Peer *et al.*, 1992, 1993; Wainright *et al.*, 1993). However, in the case of the amoebae, the morphological variation and the few ssrRNA sequences currently available indicate that amoebae are polyphyletic (Page & Blanton, 1985; Clark & Cross, 1988; Bhattacharya *et al.*, 1995). A number of taxonomic revisions have been introduced to recognise the diversity of the amoeboid form and to try to identify monophyletic groups of amoebae (Corliss, 1984; Lee *et al.*, 1985; Cavalier-Smith, 1986, 1993; Margulis *et al.*, 1989). However, the extent of the polyphyly of groups is still undetermined, and the amoebae and some of the ancient taxa which diverge early in eukaryotic history are clearly under represented in the ssrRNA phylogeny (Patterson, 1993; Hinkle & Sogin, 1993).

Ultrastructural studies recognized that the naked lobose schizopyrenids were not related to other members of the subclass Gymnamoebia (Rhizopoda, Lobosea). Initially, a new subclass was erected to accommodate the schizopyrenids (Page & Blanton, 1985), however, this subclass has since been moved from the class Lobosea to a separate phylum (Cavalier-Smith, 1993). More recently ssrRNA analysis has demonstrated that the phylum Rhizopoda is still polyphyletic because the two classes Lobosea and Filosea are not related (Bhattacharya *et al.*, 1995). The ssrRNA phylogeny presented in this study using *V. anglica* confirms these earlier findings, but also suggests that the subclass Gymnamoebia is polyphyletic. Prior to this study, ssrRNA sequence data was only available from two genera of naked lobose amoebae, *Acanthamoeba* and *Hartmannella*, from the subclass Gymnamoebia.

*Vannella anglica* did not associate with either of these genera, diverging considerably earlier in the eukaryotic tree. The exact position at which *V. anglica* branches from the eukaryotic tree is not entirely clear, but it did consistently branch at about the same point as the cellular slime mold *Dictyostelium discoideum*, before the divergence of the crown taxa which include the alveolates, stramenopiles, green algae and plants, fungi, metazoa, filose testate amoebae and the *Acanthamoeba-Hartmannella* lineages.

Considering that *Acanthamoeba* is classified in a separate order from *Vannella* and *Hartmannella* on the basis of a central filamentous core in subpseudopodia (Page, 1987), it appears that the importance of this character may be overestimated. Besides the overall shape of *Vannella* and *Platyamoeba*, there are no overwhelming characters which make these amoebae distinct from *Acanthamoeba* and *Hartmannella*. To clarify the phylogenetic relationships among the Gymnamoebia, and to identify useful phylogenetic characters for taxonomic revisions, sequence analysis of other representatives of this subclass is required. An initial survey should include at least one genus from each family from the four Gymnamoebia orders. Appropriate candidates, to supplement the sequence data from *Acanthamoeba*, *Hartmannella* and *Vannella*, could include *Chaos*, *Thecamoeba*, *Mayorella*, *Neoparamoeba*, *Gephyramoeba*, *Stereomyxa*, *Flabellula*, *Rhizamoeba*, *Corallomyxa*, *Echinoamoeba* and *Hyalodiscus*.

Looking at the amoebae from a broader perspective, it is clear that our understanding of the phylogenetic relationships among the amoeboid forms would be greatly improved with a systematic survey of ssrRNA sequences from representative well described amoebae currently in culture. There are in excess of 440 aligned eukaryotic ssrRNA sequences currently available on the ribosomal database (Maidak *et al.*, 1996), but no lobose testate amoebae (subclass Testacealobosia) or naked filose amoebae (subclass Aconchulina) have been examined from the phylum Rhizopoda. There are no ssrRNA sequences from the naked and testate reticulate amoebae (Granuloreticulosea), the intracellular plasmodial parasites (Plasmodiophorea) or other unusual amoeboid groups like the Caryoblastea and the Xenophycophorea, which together make up four phyla according to a recent classification system (Cavalier-Smith, 1993). Moreover, sequence data from the plasmodial slime molds, the cellular slime molds and the pathogenic enteric amoebae is currently available for just a single species in each

case. A survey of amoebae ssrRNA sequences would allow the genetic diversity of the amoebae to be determined with respect to other eukaryotic organisms. In combination with morphological and ultrastructural studies, a thorough ssrRNA phylogeny would provide a framework for the classification of these organisms based on monophyletic groupings, and increase the utility of molecular techniques for the identification of isolates which are difficult to culture.

Light microscopy, riboprinting, electron microscopy and sequence comparisons indicated that Species E was quite different from other naked, fan-shaped amoebae. Due to the lack of diagnostic features visible by light microscopy, few small fan-shaped amoebae, with a maximum dimension of less than 10  $\mu\text{m}$ , have been described (Sawyer, 1975b; Page, 1980a), and only one, *V. aberdonica*, has been examined by electron microscopy (Page, 1980a). This present study is the first to use both light and electron microscopy, and molecular approaches for the examination of one of these small, common and easily overlooked inhabitants of the marine benthos. Although Species E shares some morphological features with species of *Vannella* and *Platyamoeba* at the light microscopical level, the ultrastructural and molecular evidence sets it apart as an undescribed isolate. Species E may represent the first light and electron microscopical description of a *Clydonella* species. *Clydonella* possess morphological features which are intermediate between those of *Vannella* and *Platyamoeba* (Page, 1983), however, the tendency of Species E to adopt a linguiform locomotive shape, and the short irregular non-forked pseudopodia suggest that it is not a *Clydonella* species. Electron microscopical studies revealed that this isolate has a very thin undifferentiated glycocalyx, which is also characteristic of fan-shaped *Flabellula* or *Paraflabellula*, although these genera have conspicuous uroidal filaments. Moreover, despite the branched tubular mitochondrial cristae, which are typical of *Gymnamoebia*, no Golgi dictyosomes were observed which would indicate that Species E may represent a more ancient eukaryotic lineage. At this time the available evidence suggests that Species E should be assigned to a new genus. The abundance of small, unidentified, naked amoebae found in the marine benthos (Butler & Rogerson, 1995), and the unexpected results concerning Species E, indicate that this habitat at least, may host a whole assemblage of naked amoebae with a diversity of unknown proportions.

Studies have shown that small, naked, fan-shaped and limax amoebae are most abundant in benthic sediments (Sawyer, 1980; Butler, 1994; Butler & Rogerson,

1995), and comprise a significant proportion of the naked amoebae in the water column and on the surface of macroalgae (Rogerson, 1991; Rogerson & Laybourn-Parry, 1992a; Anderson & Rogerson, 1995). Consumption and growth rates also indicate that small naked amoebae may be the primary consumers of bacteria in some habitats. A fundamental examination of the identity and diversity of these small, fan-shaped and limax amoebae is essential if we are to appreciate the ecological importance of these organisms in microbial food webs.

Enrichment culture is currently the most effective method for detecting naked amoebae from field samples. However, this technique selects for morphotypes which can most easily adapt to the culture conditions, and identification generally relies on obtaining large numbers of clonal cells, which may not be possible in the laboratory. To reliably investigate the diversity of the amoebae it is therefore necessary to examine samples directly.

The *ssrRNA* has been extensively used for examining the diversity of microorganisms without the selective bias involved in cultivation and the isolation of clonal cultures (Pace *et al.*, 1985; Giovannoni *et al.*, 1990; Britschgi & Fallon, 1994; Olsen, 1994). In early studies total RNA was extracted and sequenced directly (Lane *et al.*, 1985), although this method has been superseded by DNA extraction and the selective amplification, cloning and characterisation of the target *rRNA* gene (Schmidt *et al.*, 1991). This molecular approach has been mostly utilised for the examination of eubacteria and archaeobacteria communities in a range of habitats (e.g. Ward *et al.*, 1990; Schmidt *et al.*, 1991; McInerney *et al.*, 1995), but may be applied to eukaryotic organisms to examine the diversity of naked amoebae and other protists which are difficult to detect, culture or clonally isolate.

The *ssrRNA* genes could be amplified from DNA extracted from environmental samples using universal eukaryotic primers. Amplified sequences could be cloned, characterised by restriction analysis and representative clones sequenced to develop the *ssrRNA* database of protist sequences for particular habitats. This will reveal the *ssrRNA* sequence diversity of communities of amoebae and other protists, but this will not necessarily identify the origin of each sequence if the *ssrRNA* database is poorly represented. Comparisons with other *ssrRNA* sequences may provide a broad indication of the source of the sequence, otherwise probes complementary to unique sequence elements may be required for the identification of the specific

organism in mixed samples using *in situ* hybridization.

Ribosomal RNA is an ideal candidate for the identification of micro-organisms using hybridization because the high cellular rRNA content in metabolically active cells provides abundant target sites for probe binding. The use of labelled oligonucleotide probes directed at specific regions of ssrRNA has been extensively used to identify fixed microbial cells from habitats such as soils, biofilms and activated sludge (Hahn *et al.*, 1992; Poulsen *et al.*, 1993; Wagner *et al.*, 1994). The probes are labelled with fluorescein or haptens, and hybridized cells can be detected by epifluorescence microscopy, scanning confocal laser microscopy or antibody-enzyme conjugates (Amann *et al.*, 1990; Zarda *et al.*, 1991; Hahn *et al.*, 1993; Assmus *et al.*, 1995). This methodology has been almost exclusively used for the identification of bacterial groups, but may also be applied for the identification of eukaryotic organisms (Lim, 1996). Fluorescein-labelled rRNA probes have been used to identify the ciliate *Uronema*, the flagellate *Cafeteria* and phytoplankton from natural mixed assemblages of protists (Lim *et al.*, 1993; Simon *et al.*, 1995; Knauber *et al.*, 1996). This technology would be extremely useful for the direct identification of naked amoebae among a mixed assemblage of protists in environmental samples for ecological studies. Direct counts of environmental samples could then be used to accurately determine temporal and spatial distributions of individual species or distinct phylogenetic groups of amoebae.

The complete ssrRNA gene of *V. anglica* and partial ssrRNA sequences from related amoebae has highlighted certain regions which may be useful for targeting rRNA probes. Essentially probes can be designed to hybridise with any ssrRNA variable region, but a majority of the hyper-variable regions of *V. anglica* are A+T rich, and probes complementary to these regions are likely to be non-specific. However, the E8-1 helix may be particularly useful, as this region has a balanced base composition, is absent in most eukaryotes and no known organisms have a similar sequence. Comparisons of the partial V7V8 sequences from a variety of *Vannella* and *Platyamoeba* species has revealed a sequence element in the V7 region which is potentially useful for identification of these two genera. A labelled probe directed against this sequence element could be used to distinguish *Vannella* and *Platyamoeba* species from small morphologically similar naked amoebae. Similarly, examination of the partial ssrRNA sequence of Species E ssrRNA has revealed a large insertion located within helix 43 which may be suitable for targeting ribosomal

probes for the direct identification of this isolate, and perhaps other closely related organisms.

These unique rRNA signatures could also be used to detect specific amoebae indirectly. For instance, nucleic acids can be extracted from environmental samples and the rRNA signatures characteristic of particular species can be detected on filters or slides by hybridization with labelled rRNA-targeted probes. Alternatively, DNA oligonucleotide primers complementary to unique rRNA sequence elements could be used for selective amplification in PCR assays (Vodkin *et al.*, 1992; Johnson *et al.*, 1993; Lim *et al.*, 1993; Lim, 1996). The use of indirect methods for detection of organisms are far simpler than direct methods though not without potential pitfalls. Due to the nucleic acid extraction procedure, sensitivity may be impaired and estimating numbers of target organisms in the sample may be difficult to derive.

In conclusion, this study has shown the potential of the small-subunit ribosomal RNA for identification and phylogenetic analysis of naked amoebae. Restriction analysis of the *ssrRNA* gene from morphologically similar isolates of fan-shaped amoebae can be used to distinguish between different species for identification purposes. Sequence analysis of the gene has revealed sites which may be used for targeting rRNA probes for the identification of naked amoebae in field samples containing a mixed assemblage of protists for ecological studies. Partial sequence analysis may also be used to examine the phylogenetic relationships between morphologically similar isolates and to assess the validity of taxonomic characters and divisions. Complete sequence analysis revealed that the naked fan-shaped amoebae represent another separate amoeboid lineage as they are not related to other naked lobose amoebae for which *ssrRNA* sequence data is available. However, the results of this thesis represent only a small step in the development of our knowledge of the phylogenetic relationships among amoebae. Amoebae are still very poorly represented in the *ssrRNA* database, and considering the lack of phylogenetic characteristics available, it is clear that our understanding of the evolution of these organisms is not likely to improve until a concerted effort is made to examine the *ssrRNA* sequence diversity of these organisms. In unison with morphological and ultrastructural information, *ssrRNA* sequence data will make available molecular methods for identification and ecological studies, and will aid the development of a classification scheme based upon phylogenetic relationships.

## References

- Adams, M., Andrews, R.H., Robinson, B., Christy, P., Baverstock, P.R. & Dobson, P.J. (1989). A genetic approach to species criteria in the amoeba genus *Acanthamoeba* using allozyme electrophoresis. *International Journal of Parasitology* 19:823-834.
- Amann, R. L., Krumholz, L. & Stahl, D.A. (1990). Fluorescence oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. *Journal of Bacteriology* 172:762-770.
- Amann, R., Springer, N., Ludwig, W., Görtz, H.-D. & Schleifer, K.-H. (1991). Identification *in situ* and phylogeny of uncultured bacterial symbionts. *Nature* 351:161-164.
- Anderson, G.R. & Baker, S.C. (1993). Species differentiation in the Didymozoidae (Digenea): restriction fragment length differences in internal transcribed spacer and 5.8S ribosomal DNA. *International Journal of Parasitology* 23:133-136.
- Anderson, O.R. & Rogerson, A. (1995). Annual abundances and growth potential of gymnamoebae in the Hudson Estuary with comparative data from the Firth of Clyde. *European Journal of Protistology* 31:223-233.
- Appels, R., Gerlach, W.L., Dennis, E.S., Swift, H. & Peacock, W.J. (1980). Molecular and chromosomal organisation of DNA sequences coding for the ribosomal RNAs in cereals. *Chromosoma* 78:293-311.
- Ariztia-Carmona, E.V., Andersen, R.A. & Sogin, M.L. (1991). A new phylogeny for chromophyte algae using 16S-like rRNA sequences from *Mallomonas papillosa* (Synurophyceae) and *Tribonema aequale* (Xanthophyceae). *Journal of Phycology* 27:428-436.
- Arndt, H. (1993). A critical revue of the importance of rhizopods (naked and testate amoebae) and actinopods (heliozoa) in lake plankton. *Marine Microbial Food Webs* 7:3-29.



Assmus, B., Hutzler, P., Kirchhof, G., Amann, R., Lawrence, J.R. & Hartmann, A. (1995). *In situ* localization of *Azospirillum brasilense* in the rhizosphere of wheat with fluorescently labeled, rRNA-targeted oligonucleotide probes and scanning confocal laser microscopy. *Applied and Environmental Microbiology* 61:1013-1019.

Baldock, B.M. & Sleigh, M.A. (1988). The ecology of benthic protozoa in rivers: seasonal variation in numerical abundance in fine sediment. *Archives of Hydrobiology* 3:409-421.

Bamforth, S.S. (1980). Terrestrial protozoa. *Journal of Protozoology* 27:33-36.

Barta, J.R., Jenkins, M.C. & Danforth, H.D. (1991). Evolutionary relationships of Avian *Emeria* species among other apicomplexan protozoa: monophyly of the apicomplexa is supported. *Molecular Biology and Evolution* 8:345-355.

Bentley, R.W., Harland, N.M. Leigh, J.A. & Collins, M.D. (1993). A *Staphylococcus aureus*-specific oligonucleotide probe derived from 16S rRNA gene sequences. *Letters in Applied Microbiology* 16:203-206.

Bernhard, D., Leipe, D.D., Sogin, M.L. & Schlegel, K.M. (1995). Phylogenetic relationships of the Nassulida within the phylum Ciliophora inferred from the complete small-subunit ribosomal RNA gene sequences of *Furgasonia liochmanni*, *Obertruria georgiana*, and *Pseudomicrothorax dubius*. *Journal of Eukaryotic Microbiology* 42:126-131.

Bhattacharya, D., Elwood, H.J., Goff, L.J. & Sogin, M.L. (1990). Phylogeny of *Gracilaria lemaneiformis* (Rhodophyta) based on sequence analysis of its small subunit ribosomal RNA coding region. *Journal of Phycology* 26:181-186.

Bhattacharya, D., Medlin, L., Wainright, P.O., Ariztia, E., Bibeau, C., Stickel, S.K. & Sogin, M.L. (1992). Algae containing chlorophylls a + c are paraphyletic: molecular evolutionary analysis of the chromophyta. *Evolution* 46:1801-1817.

Bhattacharya, D., Helmchen, T. & Melkonian, M. (1995). Molecular evolutionary analyses of nuclear-encoded small subunit ribosomal RNA identify an independent rhizopod lineage containing the Euglyphina and the Chlorarachniophyta. *Journal of Eukaryotic Microbiology* 42:65-69.

Bird, C.J., Rice, E.L., Murphy, C.A. & Ragan, M.A. (1992). Phylogenetic relationships in the Gracilariales (Rhodophyta) as determined by 18S rDNA sequences. *Phycologia* 31:510-522.

Bollon, A.P. (1982). Organisation of fungal ribosomal RNA genes. In: *The Cell Nucleus*, ed. Busch, H. & Rothblum, L., Academic Press, New York.

Böttger, E.C. (1989). Rapid determination of bacterial ribosomal RNA sequences by direct sequencing of enzymically amplified DNA. *FEMS Microbiology Letters* 65:171-176.

Bovee, E.C. (1965). An emendation of the amoeba genus *Flabellula* and a description of *Vannella* gen. nov. *Transactions of the American Microscopy Society* 84:217-227.

Bovee, E.C. (1970). The lobose amoebas I. A key to the suborder Conopodia Bovee and Jahn, 1966 and descriptions of thirteen new or little known *Mayorella* species. *Archiv für Protistenkunde* 112:178-227.

Bovee, E.C. & Jahn, T.L. (1966). Mechanisms of movement in taxonomy of Sarcodina. III. Orders, suborders, families and subfamilies in the superorder Lobida. *Systematic Zoology* 15:229-240.

Bovee, E.C. & Sawyer, T.K. (1979). *Marine Flora and Fauna of the North-Eastern United States. Protozoa: Amoebae*. National Oceanographic and Atmospheric Administration, Washington.

Bremer, B. (1991). Restriction data from chloroplast DNA for phylogenetic reconstruction: is there only one way of scoring? *Plant Systematics and Evolution* 175:39-54.

Britschgi, T.B. & Fallon, R.D. (1994). PCR-amplification of mixed 16S rRNA genes from an anaerobic, cyanide-degrading consortium. *FEMS Microbiology Ecology* 13:225-232.

Brosius, J., Palmer, M.L., Kennedy, P.J. & Noller, H.F. (1978), Complete nucleotide sequence of the 16S ribosomal RNA gene from *Escherichia coli*. *Proceedings of the National Academy of Science USA* 75:4801-4805.

Brown, W.M., Prager, E.M., Wang, A. & Wilson, A.C. (1982). Mitochondrial DNA sequences of primates: tempo and mode of evolution. *Journal of Molecular Evolution* 18:225-239.

Brown, S. & De Jonckheere, J.F. (1994). Identification and phylogenetic relationships of *Vahlkampfia* spp. (free-living amoebae) by riboprinting. *FEMS Microbiology Letters* 115:241-246.

Burnett, B.R. (1979). Quantitative sampling of microbiota of the deep-sea benthos-I. Sampling techniques and some data from the abyssal central North Atlantic. *Deep-Sea Research* 24:781-789.

Butler, H.G. (1994). The Ecological Importance of Marine Benthic Naked Amoebae. PhD Thesis, London.

Butler, H. & Rogerson, A. (1995). Temporal and spatial abundances of naked amoebae (gymnamoebae) in marine benthic sediments of the Clyde Sea area, Scotland. *Journal of Eukaryotic Microbiology* 42:724-730.

Butler, H. & Rogerson, A. (1996). Growth potential, production efficiency and annual production of marine benthic naked amoebae (gymnamoebae) inhabiting sediments of the Clyde Sea area, Scotland. *Aquatic Microbial Ecology* 10:123-129.

Byers, T.J., Bogler, S.A. & Burianek, L.L. (1983). Analysis of mitochondrial DNA variation as an approach to systematic relationships in the genus *Acanthamoeba*. *Journal of Protozoology* 30:198-203.

Cabot, E.L. & Beckenbach, A.T. (1989). Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. *Computational Applied Bioscience* 5:233-234.

Cann, J.P. (1981). An ultrastructural study of *Mayorella viridis* (Leidy) (Amoebida: Paramoebidae), a rhizopod containing zoochlorellae. *Archiv für Protistenkunde* 124:13-23.

Cann, J.P. (1984). The ultrastructure of *Rhizamoeba flabellata* (Goodey) comb. nov. and *Leptomyxa reticulata* Goodey (Acarpomyxia: Leptomyxida). *Archiv für Protistenkunde* 128:13-23.

Cann, J.P. & Page, F.C. (1982). Fine structure of small free-living *Paramoeba* (Amoebida) and taxonomy of the genus. *Journal of Marine Biological Association UK* 62:25-43.

Caron, D.A., Goldman, J.C. & Dennett, M.R. (1986). Enrichment of microbial populations in macroaggregates (marine snow) from surface waters of the North Atlantic. *Journal of Marine Research* 44:543-565.

Cavalier-Smith, T. (1981). Eukaryotic kingdoms - seven or nine? *Biosystems* 15:461-481.

Cavalier-Smith, T. (1986). The kingdom Chromista: origin and systematics. *Progress in Phycological Research* 5:309-347.

Cavalier-Smith, T. (1993). Kingdom Protozoa and its 18 phyla. *Microbiological Reviews* 57:953-994.

Chakrabarti, D., Dame, J.B., Gutell, R.R. & Yowell, C.A. (1992). Characterization of the rDNA unit and sequence analysis of the small subunit rRNA and 5.8S rRNA genes from *Tritrichomonas foetus*. *Molecular and Biochemical Parasitology* 52:75-84.

Chalwatzis, N. Baur, A., Stetzer, E., Kinzelbach, R. & Zimmerman, K. (1994). GenEMBL accession number X77784 (Unpublished).

- Chatton, E. (1953). Ordre des amoébiens nus ou Amoebaea. In: *Traité de Zoologie*, 1 (2), ed. Grassé P.P., Masson et Cie, Paris.
- Chávez, L.A., Balamuth, W. & Gong, T. (1986). A light- and electron-microscopical study on a new polymorphic free-living amoeba *Phreatamoeba balamuthi* n. g. n. sp. *Journal of Protozoology* 33:397-404.
- Clarholm, M. (1981). Protozoan grazing of bacteria in soil - impact and importance. *Microbial Ecology* 7:343-350.
- Clark, J.M. (1988). Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eukaryotic DNA Polymerases. *Nucleic Acids Research* 16:9677-9686.
- Clark, C.G. (1992). Riboprinting: a molecular approach to the identification and taxonomy of protozoa. In: *Protocols in Protozoology*, D4.1-D4.4, ed. Lee, J.J. & Soldo, A.T., Allen Press, Lawrence, Kansas.
- Clark, C.G. & Cross, G.A.M. (1987). rRNA genes of *Naegleria gruberi* are carried exclusively on a 14-kilobase-pair plasmid. *Molecular and Cellular Biology* 7:3027-3031.
- Clark, C.G. & Cross, G.A.M. (1988). Small-subunit ribosomal RNA sequence from *Naegleria gruberi* supports the polyphyletic origin of amoebas. *Molecular Biology and Evolution* 5:512-518.
- Clark, C.G. & Diamond, L.S. (1991a). The Larendo strain and other '*Entamoeba histolytica*-like' amoebae are *Entamoeba moshkovskii*. *Molecular and Biochemical Parasitology* 46:11-18.
- Clark, C.G. & Diamond, L.S. (1991b). Ribosomal genes of 'pathogenic' and 'non-pathogenic' *Entamoeba histolytica* are distinct. *Molecular and Biochemical Parasitology* 49:297-303.
- Clark, C.G., Cross, G.A.M. & De Jonckheere, J.F. (1989). Evaluation of evolutionary divergence in the genus *Naegleria* by analysis of ribosomal DNA plasmid restriction patterns. *Molecular and Biochemical Parasitology* 34:281-296.

Clark, C.G., Martin, D.S. & Diamond, L.S. (1995). Phylogenetic relationships among anuran trypanosomes as revealed by riboprinting. *Journal of Eukaryotic Microbiology* 42:92-96.

Cockburn, A.F., Taylor, W.C. & Firtel, R.A. (1978). *Dictyostelium* rDNA consists of non-chromosomal palindromic dimers containing 5S and 36S coding regions. *Chromosoma* 70:19-29.

Coleman, D.C. (1994). The microbial loop concept as used in terrestrial soil ecology. *Microbial Ecology* 28:245-250.

Corliss, J.O. (1979). The impact of electron microscopy on ciliate systematics. *American Zoologist* 19:573-587.

Corliss, J.O. (1984). The Kingdom Protista and its 45 phyla. *Biosystems* 17: 87-126.

Costas, M. & Griffiths, A.J. (1980). The suitability of starch-gel electrophoresis of esterases and acid-phosphatases for the study of *Acanthamoeba* taxonomy. *Archiv für Protistenkunde* 123:272-279.

Costas, M. & Griffiths, A.J. (1984). The esterases and acid-phosphatases of *Acanthamoeba* (Amoebida, Acanthamoebidae). *Protistologica* 20:33-41.

Costas, M. & Griffiths, A.J. (1985). Enzyme composition and the taxonomy of *Acanthamoeba*. *Journal of Protozoology* 32:604-607.

Craxton, M. (1991). Linear PCR amplification sequencing: a powerful method for sequencing DNA. *Methods* 3:20-24.

Creasey, A., D'Angio, L., Dunne, T.S., Kissinger, C., O'Keeffe, T., Perry-O'Keeffe, H., Moran, L.S., Roskey, M., Schildkraut, I., Sears, L.E. & Slatko, B. (1991). Application of a novel chemiluminescence based DNA detection method to single-vector and multiplex DNA sequencing. *BioTechnology* 11:102-109.

D'Alessio, J.M., Harris, G.H., Perna, P.J. & Paule, M.R. (1983). Ribosomal ribonucleic acid repeat unit of *Acanthamoeba castellanii*: cloning and restriction endonuclease map. *Biochemistry* 20:3822-3827.

Dams, E., Hendriks, L., Van de Peer, Y., Neefs, J.M. Smits, G. Van den Bempt, I. & De Wachter, R. (1988). Compilation of small subunit ribosomal-subunit RNA sequences. *Nucleic Acids Research* 16:87-173.

Davis, P.G., Caron, D.A. & Sieburth, J.McN. (1978). Oceanic amoebae from the North Atlantic: culture, distribution and taxonomy. *Transactions of the American Microscopy Society* 97:73-86.

De Jonckheere, J.F. (1983). Isoenzyme and total protein analysis by agarose isoelectric focusing, and taxonomy of the genus *Acanthamoeba*. *Journal of Protozoology* 30:701-706.

De Jonckheere, J.F. (1987). Characterization of *Naegleria* spp. by restriction endonuclease digestion of whole cell DNA. *Molecular and Biochemical Parasitology* 24:55-66.

De Jonckheere, J.F. (1988). Geographical origin and spread of pathogenic *Naegleria fowleri* from restriction enzyme patterns of repeated DNA. *BioSystems* 21:269-275.

De Jonckheere, J.H. (1993). A group I Intron in the SSUrRNA of some *Naegleria* spp. demonstrated by polymerase chain reaction amplification. *Journal of Eukaryotic Microbiology* 40:179-187.

De Jonckheere, J.F. (1994a). Riboprinting of *Naegleria* spp.: small-subunit versus large-subunit rDNA. *Parasitology Research* 80:230-234.

De Jonckheere, J.F. (1994b). Comparisons of partial ssurDNA sequences suggests revisions of species names in the genus *Naegleria*. *European Journal of Parasitology* 30:333-341.

De Jonckheere, J.F. & Brown, S. (1995). *Willaertia minor* is a species of *Naegleria*. Evidence that *Naegleria* flagellates can divide. *European Journal of Parasitology* 31:58-62.

De Lamballerie, X., Zandotti, C., Vignoli, C., Bollet, C. & De Micco, P. (1992). A one-step microbial DNA extraction method using "Chelex 100" suitable for gene amplification. *Research in Microbiology* 143:785-790.

De Long, E.F. (1992). Archaea in costal marine environments. *Proceedings of the National Academy of Science USA* 89:5685-5689.

De Rijk, P., Neefs, J-M., Van de Peer, Y. & De Wachter, R. (1992). Compilation of small ribosomal-subunit RNA sequences. *Nucleic Acids Research* 20:2075-2089.

De Rijk, P., Van de Peer, Y., Van den Broeck, I. & De Wachter, R. (1995). Evolution according to large ribosomal subunit RNA. *Journal of Molecular Evolution* 41:366-375.

De Stasio, E.A., Göringer, H.U., Tapprich, W.E. & Dahlberg, A.E. (1987). Probing ribosome function through mutagenesis of ribosomal RNA. In: *Genetics of Translation*, ed. Tuite, M.F., Picard, M. & Bolotin-Fukuhara, M., Springer-Verlag, Heidelberg.

De Stasio, E.A., Moazed, D. Noller, H.F. & Dahlberg, A.E. (1989). Mutations in the 16S ribosomal RNA disrupt antibiotic-RNA interactions. *EMBO Journal* 8:1213-1216.

Devereux, J., Haeberli, P & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* 12:387-395.

Dixon, M.T. & Hillis, D.M. (1993). Ribosomal RNA secondary structure compensatory mutations and implications for phylogenetic analysis. *Molecular Biology and Evolution* 10:256-267.

Dodge, J.D. (1973). *The Fine Structure of Algal Cells*. Academic Press, London.



Douglas, S.E., Murphy, C.A., Spencer, D.F. & Gray, M.W. (1991). Cryptomonad algae are evolutionary chimaeras of two phylogenetically distinct unicellular eukaryotes. *Nature* 350: 148-151.

Dover, G. (1982). Molecular drive: a cohesive model of species evolution. *Nature* 299:111-117

Dowling, T.E., Moritz, Palmer, J.D. & Rieseberg, L.H. (1996). Nucleic Acids III: analysis of fragments and restriction sites. In: *Molecular Systematics*, ed. Hillis, D.M., Moritz, C., Mable, B.K., Sinauer Associates Inc., Sunderland, Massachusetts.

Edwards, A.W.F. & Cavalli-Starza, L.L. (1964). Reconstruction of evolutionary trees. In: *Phenetic and Phylogenetic Classification*, ed. Heywood, V.H. & McNeil, J., Systematic Association No. 6, London.

Egger, K.N., Osmond, G. & Goodier, J.L. (1995). Sequence and putative secondary structure of group I introns in the nuclear-encoded ribosomal RNA genes of the fungus *Hymenascyphus ericase*. *Biochimica et Biophysica Acta* 1261:275-278.

Ellis, J., Hefford, C., Baverstock, P., Dalrymple, B.P. & Johnson, A. (1992). Ribosomal RNA sequence comparison of *Babesia* and *Theileria*. *Molecular and Biochemical Parasitology* 54:87-96.

Elwood, H.J., Olsen, G.J. & Sogin, M.L. (1985). The small subunit ribosomal RNA gene sequences from the hypotrichous ciliates *Oxytrichia nova* and *Stylonychia pustulata*. *Journal of Molecular Biology and Evolution* 2:399-410.

Embley, T.M. (1991). The linear PCR reaction: a simple robust method for sequencing amplified rRNA genes. *Letters in Applied Microbiology* 13:171-174.

Embley, T.M., Finlay, B.J., Thomas, R.H. & Dyal, P.L. (1992). The use of rRNA sequences and fluorescent probes to investigate the phylogenetic positions of the anaerobic ciliate *Metopus palaeformis* and its archaeobacterial endosymbiont. *Journal of General Microbiology* 138:1479-1487.

Embley, T.M., Hirt, R.P. & Williams, D.M. (1994). Biodiversity at the molecular level: the domains, kingdoms and phyla of life. *Philosophical Transactions of the Royal Society, London B* 345:21-33.

Engberg, J. (1985). The ribosomal RNA genes of *Tetrahymena*: structure and function. *European Journal of Cell Biology* 36:133-151.

Engberg, J. & Pearlman, R.E. (1972). The extrachromosomal rDNA of *Tetrahymena*. *European Journal of Biochemistry* 26:393

Erland, S., Henrion, B., Martin, F., Glover, L.A. & Alexander, I.J. (1994). Identification of the ectomycorrhizal basidiomycete *Tylospora fibrillosa* Donk by RFLP analysis of the PCR-amplified ITS and IGS region of ribosomal DNA. *New Phytologist* 126:525-532.

Eschbach, S., Wolters, J. & Sitte, P. (1991). Primary and secondary structure of the nuclear small subunit ribosomal RNA of the cryptomonad *Pyrenomonas salina* as inferred from the gene sequence: evolutionary implications. *Journal of Molecular Evolution* 32:247-252.

Fan, M., Currie, B.P., Gutell, R.R., Ragan, M.A. & Casadevall, A. (1994). The 16S-like, 5.5S-like and 23S-like ribosomal-RNAs of the 2 varieties of *Cryptococcus neoformans* - secondary structure, phylogenetic analysis and restriction fragment polymorphisms. *Journal of Medical and Veterinary Mycology* 32:163-180.

Farris, J.J. (1972). Estimating phylogenies from distance matrices. *American Naturalist* 106:645-668.

Felsenstein, J. (1978). Cases in which parsimony and compatability methods will be positively misleading. *Systematic Zoology* 27:401-410.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences. A maximum likelihood approach. *Journal of Molecular Evolution* 17:368-376.

Felsenstein, J. (1985). Confidence limits in phylogenies: an approach using the bootstrap. *Evolution* 39:783-791.

- Felsenstein, J. (1988). Phylogenies from molecular sequences: inference and reliability. *Annual Review of Genetics* 22:521-565.
- Felsenstein, J. (1993). PHYLIP - phylogeny inference package (version 3.5c). Department of Genetics, University of Washington, Seattle.
- Fenchel, T. (1987). *Ecology of Protozoa*. Science Tech Publishers, Madison, Wisconsin.
- Fenger, C.K., Granstrom, D.E., Langemeier, J.L., Gajadhar, A., Cothran, G., Tramontin, R.R., Stamper, S. & Dubey, J.P. (1994). Phylogenetic relationship of *Sarcocystis neurona* to other members of the family Sarcocystidae based on small-subunit ribosomal-RNA gene sequence. *Journal of Parasitology* 80:966-975.
- Findlay, R.C. & Gall, J.G. (1978). Free ribosomal RNA genes in *Paramecium* are tandemly repeated. *Proceedings of the National Academy of Science USA* 75:3312-3316.
- Finlay B.J., Clark, K.J., Cowling, A.J., Hindle, R.M., Rogerson, A. & Berninger, U.-G. (1988). On the abundance and distribution of protozoa and their food in a productive freshwater pond. *European Journal of Protistology* 23:205-217.
- Fitch, W.M. & Margoliash, E. (1967). Construction of phylogenetic trees. *Science* 155:279-284.
- Fitch, W.M. (1971). Toward defining the cause of evolution: minimal charge for a specific tree topology. *Systematic Zoology* 20:406-416.
- Foissner, W. (1987). Soil Protozoa: fundamental problems, ecological significance, adaptations in ciliates and testaceans, bioindicators and guide to the literature. In: *Progress in Protistology Volume 2*, ed. Corliss, J.O. & Patterson, D.J., Biopress Ltd, Bristol.

Fong, D., Man-Ying Chan, M., Rodriguez, R., Chen, C.C., Liang, Y. & Littlewood, D.T.J. (1993). Small subunit ribosomal RNA gene sequence of the parasitic protozoan *Haplosporidium nelsoni* provides a molecular probe for the oyster MSX disease. *Molecular and Biochemical Parasitology* 62:139-142.

Foolad, M.R., Jones, R.A. & Rodriguez, R.L. (1993). RAPD markers for constructing intraspecific tomato maps. *Plant Cell Reports* 12:293-297.

Fox, G.E. & Woese, C.R. (1975). 5S RNA secondary structure. *Nature* 256:505-507.

Fox, G.E., Stackbrandt, E., Hespell, R.B., Gibson, J., Maniloff, J., Dyer, T.A., Wolfe, R.S., Balch, W.E., Tanner, R.S., Magrum, L.J., Zablen, L.B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B.J., Stahl, D.A., Luehrsen, K.R., Chen, K.N. & Woese, C.R. (1980). The phylogeny of prokaryotes. *Science* 209:457-463.

Frenzel, J. (1892). Untersuchungen über die mikroskopische Fauna Argentinien. *Bibliotheca Zoologica* 4:1-82.

Fuhrman, J.A., Comeau, D.E., Hagstrom, A. & Chan, A.M. (1988). Extraction from natural planktonic microorganisms of DNA suitable for molecular biological studies. *Applied and Environmental Microbiology* 54:1426-1429.

Fuhrman, J.A., McCallum, K. & Davis, A.A. (1992). Novel major archaeobacterial group from marine plankton. *Nature* 356:148-149.

Gagnon, S., Bourbeau, D. & Levesque, R.C. (1996). Secondary structure and features of the 18S, 5.8S and 26S ribosomal RNAs from the Apicomplexa parasite *Toxoplasma gondii*. *Gene* 173:129-135.

Gajadhar, A.A., Marquardt, W.C., Hall, R., Gunderson, J., Ariztia-Carmona, E.V. & Sogin, M.L. (1991). Ribosomal RNA sequences of *Sarcocystis muris*, *Theileria annulata* and *Cryptosporidium parvum* reveal evolutionary relationships among apicomplexans, dinoflagellates and ciliates. *Molecular and Biochemical Parasitology* 45:147-154.

Gast, R.J., Fuerst, P.A. & Byers, T.J. (1994a). Discovery of group I introns in the nuclear small subunit ribosomal RNA genes of *Acanthamoeba*. *Nucleic Acids Research* 22:592-596.

Gast, R.J., Ledee, D.R., Fuerst, P.A. & Byers, T.J. (1994b). GenEMBL accession numbers U07402, U07415, U07407. (Unpublished).

Gast, R.J., Ledee, D.R., Fuerst, P.A. & Byers, T.J. (1996). Subgenus systematics of *Acanthamoeba*: four nuclear 18S rDNA sequence types. *Journal of Eukaryotic Microbiology* 43:498-504.

Gelfand, D.H. & White, T.J. (1990). Thermostable DNA polymerases. In: *PCR Protocols. A Guide to Methods and Applications*, ed. Innis, M.A., Gelfand, D.H., Sninsky, J.J. & White, T.J., Academic Press, London.

Gicquaud, C.R. (1979). Étude de l'ultrastructure du noyau et de la mitose de *Entamoeba histolytica*. *Biologie Cellulaire* 35:305-312.

Giovannoni, S.J., Britschgi, T.B., Moyer, C.L. & Field, K.G. (1990). Genetic diversity in Sargasso Sea bacterioplankton. *Nature* 345:60-63.

Goggin, C.L. (1994). Variation in the two internal transcribed spacers and 5.8S ribosomal RNA from five isolates of the marine parasite *Perkinsus* (Protista, Apicomplexa). *Molecular and Biochemical Parasitology* 65:179-182.

Goodfellow, L.P., Belcher, J.H. & Page, F.C. (1974). A light- and electron-microscopical study of *Sappinia diploidea*, a sexual amoeba. *Protistologica* 10:207-216.

Grell, K.G. (1990). Some light microscopic observations on *Chlorarachnion reptans* Geitler. *Archiv für Protistenkunde* 170:3584-3592.

Gunderson, J.H. & Sogin, M.L. (1986). Length variation in eukaryotic rRNAs: small subunit rRNAs from the protists *Acanthamoeba castellanii* and *Euglena gracilis*. *Gene* 44:63-70.

Gunderson, J.H., Elwood, H., Ingold, A., Kindle, K. & Sogin, M.L. (1987). Phylogenetic relationships between chlorophytes, chrysophytes, and oomycetes. *Proceedings of the National Academy of Science USA* 84:5823-5827.

Gunderson, J.H., Goss, S.J. & Sogin, M.L. (1994). The sequence of the *Hartmannella vermiformis* small subunit rRNA coding region. *Journal of Eukaryotic Microbiology* 41:481-482.

Gunderson, J.H., Hinkle, G., Leipe, D., Morrison, H.G., Stickel, S.K., Odelson, D.A., Breznak, J.A., Nerad, T.A., Muller, M. & Sogin, M.L. (1995). Phylogeny of trichomonads inferred from small-subunit rRNA sequences. *Journal of Eukaryotic Microbiology* 42:411-415.

Gutell, R.R. (1993). Collection of small subunit (16S- and 16S-like) ribosomal RNA structures. *Nucleic Acids Research* 13:3051-3054.

Gutell, R.R. (1994). Collection of small subunit (16S- and 16S-like) ribosomal RNA structures. *Nucleic Acids Research* 22:3502-3507.

Gutell, R.R., Weiser, B., Woese, C.R. & Noller, H.F. (1985). Comparative anatomy of 16S-like ribosomal RNA. *Progress in Nucleic Acid Research and Molecular Biology* 32:155-216.

Gutell, R.R., Noller, H.F. & Woese, C.R. (1986). Higher order structure in ribosomal RNA. *EMBO Journal* 5:1111-1113.

Gutell, R.R., Power, A., Hertz, G.Z., Putz, E.J. & Stormo, G.D. (1992). Identifying constraints on the higher-order structure of RNA: continued development and application of comparative sequence analysis methods. *Nucleic Acids Research* 20:5785-5795.

Gutell, R.R., Larsen, N. & Woese, C.R. (1994). Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiological Reviews* 58:10-26.

- Hahn, D., Amann, R.I., Ludwig, W., Akkermans, A.D.L. & Schleifer, K.-H. (1992). Detection of micro-organisms in soil after *in situ* hybridization with rRNA-targeted, fluorescently labelled oligonucleotides. *Journal of General Microbiology* 138:879-887.
- Hahn, D., Amann, R.I. & Zeyer, J. (1993). Whole-cell hybridization of *Frankia* strains with fluorescence- or dioxygenin-labeled, 16S rRNA-targeted oligonucleotide probes. *Applied and Environmental Microbiology* 59:1709-1716.
- Hardys, H., Balick, M. & Schierwater, B. (1992). Applications of RADP in molecular ecology. *Molecular Ecology* 1:55-63.
- Harrison, J., Molloy, P.L. & Clark, S.J. (1994). Direct cloning of polymerase chain reaction products in an *Xcm*I T-vector. *Analytical Biochemistry* 216:235-236.
- Hartskeerl, R.A., Schuitema, A.R.J. & De Wachter, R. (1993). Secondary structure of the small subunit ribosomal RNA sequence of the microsporidium *Encephalitozoon cuniculi*. *Nucleic Acids Research* 21:1489.
- Hasegawa, M., Hashimoto, T., Adachi, T., Iwabe, N. & Migata, T. (1993). Early branchings in the evolution of eukaryotes: ancient divergence of *Entamoeba* that lacks mitochondria revealed by protein sequence data. *Journal of Molecular Evolution* 36:380-388.
- Hemsley, A., Arnheim, N., Toney, M.D., Cortopassi, G. & Galas, D.J. (1989). A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucleic Acids Research* 17:6545-6551.
- Hendriks, L., Van de Peer, Y., Van Herck, M., Neefs, J.-M. & De Wachter, R. (1990). The 18S ribosomal RNA sequence of the sea anemone *Anemonia sulcata* and its evolutionary position among other eukaryotes. *FEBS Letters* 269:445-449.
- Hendy, M.D. & Penny, D. (1982). Branch and bound algorithms to determine minimal evolutionary trees. *Mathematical Bioscience* 59:277-290.
- Hendy, M.D. & Penny, D. (1989). A framework for the quantitative study of evolutionary trees. *Systematic Zoology* 38:297-309.

Hill, W.E., Dahlberg, A., Garrett, R.A., Moore, P.B., Schlessinger, D. & Warner, J.R. (1990). *The Ribosome: Structure, Function and Evolution*. American Society for Microbiology, Washington D.C.

Hillis, D.M., Mable, B.K., Larson, A., Davis, S.K. & Zimmer, E.A. (1996). Nucleic Acids IV: sequencing and cloning. In: *Molecular Systematics*, ed. Hillis, D.M., Moritz, C. & Mable, B.K., Sinauer Associates Inc., Sunderland, Massachusetts.

Hinkle, G., Leipe, D.D., Nerad, T.A. & Sogin, M.L. (1994). The unusually long small subunit ribosomal RNA of *Phreatamoeba balamuthi*. *Nucleic Acids Research* 22:465-469.

Hinkle, G. & Sogin, M.L. (1993). The evolution of the Vahlkampfiidae as deduced from 16S-like RNA analysis. *Journal of Eukaryotic Microbiology* 40:599-603.

Hogue, M.J. (1914). Studies on the life history of an amoeba of the limax group *Vahlkampfia calkinsi*. *Archiv für Protistenkunde* 35:154-163.

Hogue, M.J. (1921). Studies on the life history of *Vahlkampfia patuxent* n. sp., parasitic in the oyster, with experiments regarding its pathogenicity. *American Journal of Hygiene* 1:321-345.

Holley, R.W., Apgar, J., Everett, G.A., Madison, J.T., Marguisee, M., Merrill, S.H., Penswick, J.R. & Zamir, A. (1965). Structure of a ribonucleic acid. *Science* 147:1462-1465.

Holton, T.A. & Graham, M.W. (1990). A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. *Nucleic Acids Research* 19:1156.

Hori, H. & Osawa, S. (1986). Evolutionary change in 5S RNA secondary structure and a phylogenetic tree of 352 5S RNA species. *BioSystems* 19:163-172.

Innis, M.A. Myambo, K.B., Gelfand, D.H. & Brow, M.A.D. (1988). DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. *Proceedings of the National Academy of Science USA* 86:9436-9440.



Innis, M.A. & Gelfand, D.A. (1990). Optimization of PCRs. In: PCR protocols: A Guide to Methods and Applications, ed. Innis, M.A., Gelfand, D.H., Sninsky, J.J. & White, T.J., Academic Press, San Diego.

Innis, M.A., Gelfand, D.H., Sninsky, J.J. & White, T.J. (1990). PCR protocols: A Guide to Methods and Applications. Academic Press, San Diego.

Irvine, D.E.G. & John, D.M. (1984). Systematics of the green algae. Academic Press, London.

Iwabe, N., Kuma, K., Kishino, H., Hasegawa, M. & Miyata, T. (1991). Evolution of RNA polymerases and branching patterns of the 3 major groups of archeabacteria. *Journal of Molecular Evolution* 32:70-78.

Jahn, T.L., Bovee, E.C. & Griffith, D.L. (1974). Taxonomy and evolution of the Sarcodina: a reclassification. *Taxon* 23:483-496.

James, S.A., Collins, M.D. & Roberts, I.N. (1994). Genetic interrelationships among species of the genus *Zygosaccharyomyces* as revealed by small subunit rRNA gene sequences. *Yeast* 10:871-884.

Jeffreys, A.J., Wilson, V. & Thein, S.L. (1985). Individual-specific fingerprints in human DNA. *Nature* 316:76-79.

Johansen, T., Johansen, S. & Haugli, F.B. (1988). Nucleotide sequence of the *Physarum polycephalum* small subunit ribosomal RNA as inferred from the gene sequence: secondary structure and evolutionary implications. *Current Genetics* 14:265-273.

Johnson, A.M., Illana, S., Hakendorf, P. & Baverstock, P.R. (1988). Phylogenetic relationships of the apicomplexan protist *Sarcocystis* as determined by small-subunit ribosomal RNA comparisons. *Journal of Parasitology* 74:847-860.

Johnson, A.M., Fielke, R., Christy, P.E., Robinson, B. & Baverstock, P.R. (1990). Small subunit ribosomal RNA evolution in the genus *Acanthamoeba*. *Journal of General Microbiology* 136:1689-1698.

Johnson, D.W., Pieniazek, N.J. & Rose, J.B. (1993). DNA probe hybridization and PCR detection of *Cryptosporidium* compared to immunofluorescence assay. *Water Science and Technology* 27:77-84.

Johnston, D.A., Kane, R.A. & Rollinson, D. (1993). Small sbunit (18S) ribosomal RNA gene divergence in the genus *Schistosoma*. *Parasitology* 107:147-156.

Jung, V., Pestka, S.B. & Pestka, S. (1990). Efficient cloning of PCR generated DNA containing terminal restriction sites. *Nucleic Acids Research* 18:6156.

Karrer, K.M. & Gall, J.G. (1976). The macronuclear ribosomal DNA of *Tetrahymena pyriformis* is a palindrome. *Journal of Molecular Biology* 105:421-453.

Kaufman, D.L. & Evans, G.A. (1990). Restriction endonuclease cleavage at the termini of PCR products. *BioTechniques* 9:304-306.

Kenchington, E., Naidu, K.S., Roddick, D.L., Cook, D.I. & Zouros, E. (1993). The use of biochemical genetic markers to discriminate between adductor muscles of the sea scallop (*Placopecten magellanicus*) and the Iceland scallop (*Chlamys islandica*). *Canadian Journal of Fisheries and Aquatic Sciences* 50:1222-1228.

Kilvington, S., Beeching, J.R. & White, D.G. (1991). Differentiation of *Acanthamoeba* strains from infected corneas and the environment by using restriction endonuclease digestion of whole-cell DNA. *Journal of Clinical Microbiology* 29:310-314.

Kim, S.-H. (1979). Crystal structure of yeast tRNA-phe and general structural features of other tRNAs. In: *Transfer RNA: Structure, Properties, and Recognition*, ed. Schimmel, P.R., Soll, D. & Abelson, J.N., Cold Spring Harbor, New York.

Knauber, D.C., Berry, E.S. & Fawley, M.W. (1996). Ribosomal RNA-baed oligonucleotide probes to identify marine green ultraphytoplankton. *Journal of Eukaryotic Microbiology* 43:89-94.

Knoll, M. (1992). The early evolution of the eukaryotes: a geological perspective. *Science* 256: 622-627.

Kobayashi, M., Takahashi, M., Wada, H. & Satoh, N. (1993). Molecular phylogeny inferred from sequences of small-subunit ribosomal RNA supports the monophyly of the metazoa. *Zoological Science* 10:827-833.

Kohn, L.M., Petsche, D.M., Bailey, S.R., Novak, L.A. & Anderson, J.B. (1988). Restriction fragment length polymorphisms in nuclear and mitochondrial DNA of *Sclerotinia* species. *Phytopathology* 78:1047-1051.

Kovalic, D., Kwak, J. & Weisblum, B. (1991). General method for direct cloning of DNA fragments generated by the polymerase chain reaction. *Nucleic Acids Research* 19:4560.

Kuma, K.-I., Nikoh, N., Iwabe, N. & Miyata, T. (1995). Phylogenetic position of *Dictyostelium* inferred from multiple protein datasets. *Journal of Molecular Evolution* 41:238-246.

Kwon, O.-Y., Ogino, K. & Ishikawa, H. (1991). The longest 18S ribosomal RNA ever known. Nucleotide sequence and presumed secondary structure of the 18S rRNA of the pea aphid, *Acyrtosiphon pisum*. *European Journal of Biochemistry* 202:827-833.

Lane, D.J., Pace, B., Olsen, G.J., Stahl, D.A., Sogin, M.L. & Pace, N.R. (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Science USA* 82:6955-6959.

Lake, J.A. (1987a). Rate-independent technique for analysis of nucleic acid sequences: evolutionary parsimony. *Molecular Biology and Evolution* 4:167-191.

Lake, J.A. (1987b). Prokaryotes and archaebacteria are not monophyletic: rate invariant analysis of rRNA genes indicates that eukaryotes and eocytes form a monophyletic taxon. *Cold Spring Harbor Symposium on Quantitative Biology* 52:839-846.

Lake, J.A. (1988). Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. *Nature* 331:184-186.

Laybourn-Parry, J. & Rogerson, A. (1993). Seasonal patterns of protozooplankton in Lake Windermere, England. *Archiv für Hydrobiologie* 129:25-43.

Laybourn-Parry, J., Marchant, H.J. & Brown, P. (1991). The plankton of a large oligotrophic freshwater antarctic lake. *Journal of Plankton Research* 13:1137-1149.

Lee, J.J., Hunter, S.H. & Bovee, E.C. (1985). *An Illustrated Guide to the Protozoa*. Society of Protozoologists, Lawrence, Kan.

Leipe, D.D., Gunderson, J.H., Neras, T.A. & Sogin, M.L. (1993). Small subunit ribosomal RNA of *Hexamita inflata* and the quest for the first branch in the eukaryotic tree. *Molecular and Biochemical Parasitology* 59:41-48.

Leipe, D.D., Wainright, P.O., Gunderson, J.H., Porter, D., Patterson, D.J., Valois, F., Himmerich, S. & Sogin, M.L. (1994). The stramenopiles from a molecular perspective: 16S-like rRNA sequences from *Labyrinthuloides minuta* and *Cafeteria roenbergensis*. *Phycologica* 33:369-377.

Leonardi, M.C., Bolzern, A.M. & Fascio, U. (1991). Analisi densitometrica di patterns elettroforetici di proteine totali in amebe. *Ist Lombardo Rend Sci B* 122:85-90.

Leonardi, M.C., Bolzern, A.M., De Bernardi, F. & Fascio, U. (1995). The differentiation of free-living amoebae of the genus *Mayorella*: a biochemical approach. *Bollettino di Zoologia* 62:91-95.

Levine, N.D., Corliss, J.O., Cox, F.E.G., Deroux, G., Grain, J., Honigberg, B.M., Leedale, G.F., Loeblich III, A.R., Lom, J., Lynn, D., Merinfeld, E.G., Page, F.C., Poljansky, G., Sprague, J., Vavra, J. & Wallace, F.G. (1980). A newly revised classification of the Protozoa. *Journal of Protozoology* 27:37-58.

Levitt, M. (1969). Detailed molecular model for transfer ribonucleic acid. *Nature* 224:759-763.

Lim, E.L. (1996). Molecular identification of nanplanktonic protists based on small subunit ribosomal RNA gene sequences for ecological studies. *Journal of Eukaryotic Microbiology* 43:101-106.

Lim, E.L., Amaral, L.A., Caron, D.A. & DeLong, E.F. (1993). Application of rRNA-based probes for observing marine nanoplanktonic protists. *Applied and Environmental Microbiology* 59:1647-1655.

Lipps, H.J. & Steinbrück, G. (1978). Free genes for rRNA in the macronuclear genome of the ciliate *Stylonychia mytilus*. *Nucleic Acids Research* 5:4699-4709.

Lobet, Y., Peacock, M.G. & Cieplak Jr, W. (1989). Frame-shift mutations in the *lacZ* gene of certain commercially available pUC18 plasmids. *Nucleic Acids Research* 17:4897.

Loeffelholz, M.J. & Scholl, D.R. (1989). Method for improved extraction of DNA from *Nocardia asteroides*. *Journal of Clinical Microbiology* 27:1880-1881.

Long, M., Ding, Y., Polin, D. & Luft, B.J. (1993). GenEMBL accession number X68523. (Unpublished).

Loomis, W.F. & Smith, D. (1990). Molecular phylogeny of *Dictyostelium discoideum* by protein sequences. *Proceedings of the National Academy of Science USA* 87:9093-9097.

Lumb, S.M., Bray, R.A. & Rollinson, D. (1993). Partial small subunit (18S) rRNA gene sequences from fish parasites of the families Lepocreadiidae and Fellodistamidae (Digenea) and their use in phylogenetic analyses. *Systematic Parasitology* 26:141-149.

Lumsden, L. & Hall, D.O. (1975). Superoxide dismutase in photosynthetic organisms provides an evolutionary hypothesis. *Nature* 257:670-672.

Maddison, W.P. & Maddison, D.R. (1992). MacClade, version 3.0. Sinauer, Sunderland, Massachusetts.

- Maidak, B.L., Olsen, G.J., Larsen, N., Overbeek, R., McCaughey, M.L. & Woese, C.R. (1996). The ribosomal database project (RDP). *Nucleic Acids Research* 24:82-85.
- Makhlin, E.E. (1981). Pattern of DNA synthesis in nuclei of feeding and starving *Amoeba proteus*: a cytofluorometric study. *Experimental Cell Research* 136:417-422.
- Mandal, R.K. (1984). The organisation and transcription of eukaryotic ribosomal RNA genes. *Progress in Nucleic Acid Research* 31:115-160.
- Marché, S. Roth, C., Philippe, H., Dollet, M. & Baltz, T. (1995). Characterization and detection of plant trypanosomatids by sequence analysis of the small subunit ribosomal RNA gene. *Molecular and Biochemical Parasitology* 71:15-26.
- Marchuk, D., Drumm, M., Saulino, A. & Collins, F.S. (1990). Construction of T-vectors, a rapid and general system for direct cloning of modified PCR products. *Nucleic Acids Research* 19:1154.
- Margulis, L. & Schwartz, K.V. (1982). Five kingdoms. An illustrated guide to the phyla of life on earth. Freeman, San Fransisco.
- Margulis, L., Corliss, J.O., Melkonian, M. & Chapman, D.J. (1989). Handbook of Protoctista. Jones and Bartlett Publishers, Boston.
- Martin, G.W., Alexopoulos, C.J. & Farr, M.L. (1983). The Genera of Myxomycetes. Univeristy of Iowa Press, Iowa City.
- Maslov, D.A. Lukes, J., Jirkey, M. & Simpson, L. (1996). Phylogeny of the trypanosomes as inferred from small and large subunit rRNAs: implications for the evolution of parasitism in the trypanosomatid protozoa. *Molecular and Biochemical Parasitology* 75:197-205.
- Mayes, F.D. (1995). The Ecological Importance of Antarctic Marine Naked Amoebae. PhD Thesis, University of London.
- Maxam, A.M. & Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Science USA* 74:560-564.

- Maxam, A.M. & Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods in Enzymology* 65:499-559.
- McCallum, F.S. & Maden, B.S.H. (1985). Human 18S ribosomal RNA sequence inferred from DNA sequence. *Biochemical Journal* 232:725-733.
- McCarroll, R., Olsen, G.J., Stahl, Y.D., Woese, C.R. & Sogin, M.L. (1983). Nucleotide sequence of the *Dictyostelium discoideum* small-subunit ribosomal ribonucleic acid inferred from the gene sequence: evolutionary implications. *Biochemistry* 22:5858-5868.
- McCutchan, T.F., De la Cruz, V.F., Lal, A.A., Gunderson, J.H., Elwood, H.J. & Sogin, M.L. (1988). Primary sequences of two small subunit ribosomal RNA genes from *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* 28:63-68.
- McFadden, G.I., Gilson, P.R., Hofmann, C.J.B., Adcock, G.J. & Maier, U.G. (1992). GenEMBL accession number X70809. (Unpublished).
- McFadden, G.I., Gilson, P.R., Hofmann, C.J.B., Adcock, G.J. & Meier, U.G. (1994a). Evidence that an amoeba acquired a chloroplast by retaining part of an engulfed eukaryotic alga. *Proceedings of the National Academy of Science USA* 91:3690-3694.
- McFadden, G.I., Gilson, P.R. & Hill, D.R.A. (1994b). *Goniomonas* rRNA sequences indicate that this phagotrophic flagellate is a close relative of the host component of cryptomonads. *European Journal of Phycology* 29:29-32.
- McInerney, J.O., Wilkinson, M., Patching, J.W., Embley, T.M. & Powell, R. (1995). Recovery and phylogenetic analysis of novel archeal rRNA sequences from the deep-sea deposit feeder. *Applied and Environmental Microbiology* 61:1646-1648.
- McLaughlin, G.L., Brandt, F.H. & Visvesvara, G.S. (1988). Restriction fragment length polymorphisms of the DNA of selected *Naegleria* and *Acanthamoeba* amoebae. *Journal of Clinical Microbiology* 26:1655-1658.
- McNally, K.C., Govind, N.S., Thome, P.E. & Trench, R.K. (1993). GenEMBL accession number M88518. (Unpublished).

- Mead, D.A., Pey, N.Y., Herrnstadt, C., Marcil, R.A. & Smith, L.M. (1991). A universal method for the direct cloning of PCR amplified nucleic acid. *BioTechnology* 9:657-663.
- Medlin, L., Elwood, H.J., Stickel, S. & Sogin, M.L. (1988). The characterization of enzymatically amplified eukaryotic 16S-like rRNA coding. *Gene* 71:491-499.
- Medlin, L.K., Lange, M. & Baumann, M.E.M. (1994). Genetic evidence among three colony-forming species of *Phaeocystis*: further evidence for the phylogeny of the Prymnesiophyta. *Phycologia* 33:199-212.
- Messing, J., Carlson, J., Hagen, G., Rubenstein, I. & Oleson, A. (1984). Cloning and sequencing of the ribosomal RNA genes in maize: the 17S region. *DNA* 3:31-40.
- Minassian, I. & Bell, L.G.E. (1976). Timing of nucleolar DNA replication in *Amoeba proteus*. *Journal of Cell Science* 22:521-530.
- Moazed, D. & Noller, H.F. (1987). Interaction of antibiotics with functional sites in 16S ribosomal RNA. *Nature* 327:389-394.
- Mole, S.E., Iggo, R.D. & Lane, D.P. (1989). Using the polymerase chain reaction to modify expression plasmids for epitope mapping. *Nucleic Acids Research* 17:3319.
- Morgon, E.A., Gregory, S.T., Sigmund, C.D. & Borden, A. (1987). Antibiotic resistance mutations in *Escherichia coli* ribosomal RNA genes and their uses. In: *Genetics of Translation*, ed. Tuite, M.F., Picard, M. & Bolotin-Fukuhara, M., Springer-Verlag, Heidelberg.
- Moss, D.M., Brandt, F.H., Matthews, H.M. & Visvesvara, G.S. (1988). High resolution polyacrylamide gradient gel electrophoresis (PGGE) of isozymes from five *Naegleria* species. *Journal of Protozoology* 35:26-31.
- Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. & Erlich, H. (1986). Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harbor Symposium of Quantitative Biology* 51:263-273.



Murray, V. (1989). Improved double-stranded DNA sequencing using the linear polymerase chain reaction. *Nucleic Acids Research* 17:8889.

Neefs, J.-M. & De Wachter, R. (1990). A proposal for the secondary structure of a variable area of eukaryotic small subunit ribosomal subunit RNA involving the existence of a pseudoknot. *Nucleic Acids Research* 18:5695-5704.

Neefs, J.-M., Van de Peer, Y., De Rijk, P., Goris, A. & De Wachter, R. (1991). Compilation of small ribosomal RNA sequences. *Nucleic Acids Research* 19:1987-2015.

Neefs, J.-M. Van de Peer, Y., De Rijk, P., Chapelle, S. & De Wachter, R. (1993). Compilation of small ribosomal subunit RNA structures. *Nucleic Acids Research* 21:3025-3049.

Nerad, T.A. & Daggett, P.M. (1979). Starch gel electrophoresis: an effective method for separation of pathogenic and non-pathogenic *Naegleria* strains. *Journal of Protozoology* 26:613-615.

Nickrent, D.L. & Sargent, M.L. (1991). An overview of the secondary structure of the V4 region of eukaryotic small-subunit ribosomal RNA. *Nucleic Acids Research* 19:227-235.

Noller, H.F. (1984). Structure of ribosomal RNA. *Annual Review of Biochemistry* 53:119-162.

Noller, H.F. & Woese, C.R. (1981). Secondary structure of 16S ribosomal RNA. *Science* 212:403-411.

Noller, H.F., Kop, J., Wheaton, V., Brosius, J., Gutell, R.R., Kopylov, A.M., Dohme, F., Herr, W., Stahl, D.A., Gupta, R. & Woese, C.R. (1981). Secondary structure model for 23S ribosomal RNA. *Nucleic Acids Research* 21:6167-6189.

Olive, L.S. (1975). *The Mycetozoans*. Academic Press, New York.

Olsen, G.J. (1994). Archea, Archea, everywhere. *Nature* 371:657-658.

Olsen, G.J. & Woese, C.R. (1993). Ribosomal RNA: a key to phylogeny. *FASEB Journal* 7:113-123.

Olsen, G.J., Lane, D.J., Giovannoni, S.J., Pace, N.R. & Stahl, D.A. (1986). Microbial ecology and evolution: a ribosomal RNA approach. *Annual Review of Microbiology* 40:337-365.

Olsen, G.J., Matsuda, I.J., Hagstrom, R. & Overbeek, R. (1994). fastDNAm1: a tool for construction of phylogenetic trees of DNA sequences using the maximum likelihood. *Computer Applications in the Biosciences* 10:41-48.

Owen, R.J. & Borman, P. (1987). A rapid biochemical method for purifying high molecular weight bacterial chromosomal DNA for restriction analysis. *Nucleic Acids Research* 15:3631.

Pace, N.R., Stahl, D.A., Lane, D.J. & Olsen, G.J. (1985). Analyzing natural microbial populations by rRNA sequences. *American Society of Microbiology News* 51:4-12.

Page, F.C. (1967). Taxonomic criteria for limax amoebae, with descriptions of 3 new species of *Hartmannella* and 3 of *Vahlkampfia*. *Journal of Protozoology* 14:709-724.

Page, F.C. (1968). Generic criteria for *Flabellula*, *Rugipes*, and *Hyalodiscus*, with descriptions of species. *Journal of Protozoology* 15:9-26.

Page, F.C. (1969). *Platyamoeba stenopodia* n. g., n. sp., a freshwater amoeba. *Journal of Protozoology* 14:499-521.

Page, F.C. (1971). A comparative study of five freshwater and marine species of Thecamoebidae. *Transactions of the American Microscopy Society* 90:157-173.

Page, F.C. (1974a). Some marine *Platyamoeba* of East Anglia. *Journal of the Marine Biological Association UK* 54:651-664.

Page, F.C. (1974b). A further study of taxonomic criteria for limax amoebae, with descriptions of new species and a key to genera. *Archiv für Protistenkunde* 116:149-184.

- Page, F.C. (1976). A revised classification of the Gymnamoebia (Protozoa: Sarcodina). *Zoological Journal of the Linnean Society* 58:61-77.
- Page, F.C. (1977). The genus *Thecamoeba* (Protozoa, Gymnamoebia). Species distinctions, locomotive morphology, and protozoan prey. *Journal of Natural History* 11:25-63.
- Page, F.C. (1979a). Two genera of marine amoebae (Gymnamoebia) with distinct surface structures: *Vannella* Bovee, 1965, and *Pseudoparamoeba* n. gen, with two new species of *Vannella*. *Protistologica* 15:245-247.
- Page, F.C. (1979b). *Vexillifera armata* n. sp (Gymnamoebia, Paramoebidae), an estuarine amoeba with distinctive surface structures and trichocyst-like bodies. *Protistologica* 16:605-612.
- Page, F.C. (1980a). A key to marine species of *Vannella* (Sarcodina: Gymnamoebia), with descriptions of new species. *Journal of the Marine Biological Association UK* 60:929-946.
- Page, F.C. (1980b). Fine structure of some strains of *Platyamoeba* (Gymnamoebia, Thecamoebidae). *Protistologica* 16:605-612.
- Page, F.C. (1980c). A light and electron-microscopical comparison of marine limax and flabellate amoebae belonging to four genera. *Protistologica* 16:57-78.
- Page, F.C. (1983). Marine gymnamoebae. Institute of Freshwater Ecology, NERC, CCAP, Cambridge.
- Page, F.C. (1987). The classification of the naked amoebae (Phylum Rhizopoda). *Archiv für Protistenkunde* 133:199-217.
- Page, F.C. (1988). A new key to freshwater and soil gymnamoebae. Freshwater Biological Association, Ambleside, Cumbria.

Page, F.C. & Blakey, S.M. (1979). Cell surface structure as a taxonomic character in the Thecamoebidae (Protozoa: Gymnamoebia). *Zoological Journal of the Linnean Society* 66:113-135.

Page, F.C. & Blanton, R.L. (1985). The Heterolobosea (Sarcodina: Rhizopoda), a new class uniting the Schizopyrenida and the Acrasidae (Acrasida). *Protistologica* 21:121-132.

Page, F.C. & Siemensma, F.J. (1991) Nackte Rhizopoda and Heliozoa. In: *Protozoenfauna*, ed. Matthes, D., Gustav fischer Verlag, Stuttgart.

Page, F.C. & Willumsen, N.B.S. (1983). A light- and electron-microscopical study of *Paraflabellula reniformis* (Schmoller, 1964), type species of a genus of amoebae (Amoebidae, Flabellulidae) with subpseudopodia. *Protistologica* 19:567-575.

Patterson, D.J (1989). Stramenopiles: chromophytes from a protistan perspective. In: *The Chromophyte Algae Problems and Perspectives*, ed. Green, J.C., Leadbeater, B.S.C. & Diver, W.L., Clarendon Press, Oxford.

Patterson, D.J. (1993) Protozoa: evolution and systematics. In: *Progress in Protozoology. Proceedings of the IX International Congress of Protozoology*, Berlin, ed. Hausmann, K. & Hülsmann, M., Gustav Fischer Verlag, Stuttgart.

Pattersen, D.J. & Larsen, J. (1991). *The Biology of Free-Living Heterotrophic Flagellates*. Oxford University Press, Oxford.

Patterson, D.J. & Sogin, M.L. (1993). Eukaryotic origins and protistan diversity. In: *The Origin and Evolution of Prokaryotic and Eukaryotic Cells*, ed. Hartman, H. & Matsuno, K., World Scientific Publishing Co., New Jersey.

Pennick, N.C. & Goodfellow, L.P. (1975). Some observations on the cells surface structures of species of *Mayorella* and *Paramoeba*. *Archiv für Protistenkunde* 117:41-46.

Pernin, P., Cariou, M-L. & Jacquier, A. (1985). Biochemical identification and phylogenetic relationships in free-living amoebas of the genus *Naegleria*. *Journal of Protozoology* 32:592-603.

Pieniazek, N.J., Arrowood, M.J., Blagburn, B.L., Mathews, H.M. & Slemenda, S.B. (1993a). GenEMBL accession number L16997. (Unpublished).

Pieniazek, N.J., Moura, H., Slemenda, S.B. & Visvesvara, G.S. (1993b). GenEMBL accession number L19070. (Unpublished).

Pikaart, M.J. & Villeponteau, B. (1993). Suppression of PCR amplification by high levels of RNA. *Biotechniques* 14:24-25.

Pitcher, D., Saunders, N.A. & Owen, R.J. (1989). Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. *Letters in Applied Microbiology* 8:151-156.

Poot, R.A., Cornelis, W., Pleij, A. & Van Duin, J. (1996). The central pseudoknot in 16S ribosomal RNA is needed for ribosome stability but is not essential for 30S initiation complex formation. *Nucleic Acids Research* 24:3670-3676.

Powers, T. & Noller, H.F. (1991). A functional pseudoknot in 16S ribosomal RNA. *Nucleic Acids Research* 9:6167-6189.

Poulsen, L.K., Ballard, G. & Stahl, D.A. (1993). Use of rRNA fluorescence *in situ* hybridization for measuring the activity of single cells in young and established biofilms. *Applied and Environmental Microbiology* 59:1354-1360.

Pühler, G., Leffers, H., Gropp, F., Paln, P., Klenk, H.P., Lottspeich, F., Garrett, R.A. & Zillig, W. (1989). Archeobacterial DNA dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. *Proceedings of the National Academy of Science USA* 86:4569-4573.

Pussard, M. (1973). Description d'une amibe de type flabellulien: *Personella marginata* n. g., n. sp. (Mayorellidae, Amoebeae). *Protistologica* 9:175-185.

Pussard, M. & Pons, R. (1977). Morphologie de la paroi kystique et taxonomie du genre *Acanthamoeba* (Protozoa, Amoebida). *Protistologica* 13:557-598.

Que, X. & Reed, S. (1991). Nucleotide sequence of a small subunit ribosomal RNA (16S-like rRNA) gene from *Entamoeba histolytica*: differentiation of pathogenic from nonpathogenic isolates. *Nucleic Acids Research* 19: 5438-5438.

Ragan, M.A. & Chapman, D.J. (1978). *A Biochemical Phylogeny of the Protists*. Academic Press, New York.

Ramirez, E., Warren, A., Rivera, R., Bonilla, P., Rodriguez, S., Calderón, A., Oritz, R. & Gallegos, E. (1993). An investigation of the pathogenic and non-pathogenic free-living amoebae in an activated-sludge plant. *Water, Air and Soil Pollution* 69:93-98.

Raué, H.A., Musters, W., Rutgers, C.A., Riet, J.V. & Planta, R.J. (1990). rRNA: from structure to function. In: *The Ribosome: Structure, Function and Evolution*, ed. Hill, W.E., Dahlberg, A., Garrett, R.A., Moore, P.B., Schlessinger, D. & Warner, J.R., American Society for Microbiology, Washington D.C.

Raynal, F., Michot, B. & Bachellerie, J.P. (1984). Complete nucleotide sequence of mouse 18S rRNA gene: comparison with other available homologs. *FEBS Letters* 167:263-268.

Reddy, G.R., Chakrabarti, D., Yowell, C.A & Dame, J.B. (1991). Sequence microheterogeneity of the three small subunit ribosomal RNA genes of *Babesia bigemina*: expression in erythrocyte culture. *Nucleic Acids Research* 19:3641-3645.

Reynolds, E.S. (1963). The use of lead citrate at high pH as an electron-opaque stain in electron microscopy. *Journal of Cell Biology* 17:208-212.

Rivera, f., Rodriguez, S., Warren, A., Bonilla, P., Rodriguez, S., Calderón, A. & Oritz, R. (1993). An investigation of the pathogenic and non-pathogenic free-living amoebae from the root zone method of wastewater treatment. *Water, Air and Soil Pollution* 69:135-139.

Rogerson, A. (1982). An estimate of the annual production and energy flow of the large naked amoebae population inhabiting *Sphagnum* bog. *Archiv für Protistenkunde* 135:289-298.

Rogerson, A. (1991). On the abundance of marine naked amoebae on the surfaces of five species of macroalgae. *FEMS Microbiology Ecology* 85:301-312.

Rogerson, A. & Laybourn-Parry, J. (1992a). The abundance of marine naked amoebae in the water column of the Clyde Estuary. *Estuarine, Coastal and Shelf Science* 34:187-196.

Rogerson, A. & Laybourn-Parry, J. (1992b). Aggregate dwelling protozooplankton communities in estuaries. *Archiv für Hydrobiologie* 125:411-422.

Rowan, R. & Powers, D.A. (1992). Ribosomal RNA sequences and the diversity of symbiotic dinoflagellates (zooxanthellae). *Proceedings of the National Academy of Science USA* 89:3639-3643.

Rubtsov, P.M., Musakhanov, M.M., Zakharyev, V.M., Krayev, A.S., Skryabin, K.G. & Bayev, A.A. (1980). The structure of the yeast ribosomal RNA genes. I. The complete nucleotide sequence of the 18S ribosomal RNA gene from *Saccharomyces cerevisiae*. *Nucleic Acids Research* 8:5779-5794.

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. & Arnheim, N. (1985). Enzymatic amplification of  $\beta$ -globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230:1350-1354.

Saiki, R.K., Gelfand, D.H., Soffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., Erlich, H.A. (1988). Primer-directed enzymic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487-491.

Saitou, N. & Nei, M. (1987). The neighbour-joining method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406-425.

Sambrook, J., Fritsch, E.F., Maniatis, T. (1989). *Molecular Cloning*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sanger, F., Nicklen, S. & Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Science USA* 74:5463-5467.

Sankoff, D. (1975). Minimal mutation trees of sequences. *SIAM Journal of Applied Mathematics* 28:35-42.

Sargeant, P.G., Williams, J.E., Jones, D.M. (1982). Electrophoretic isoenzyme patterns of *Entamoeba histolytica* and *Entamoeba chattoni* in a Primate survey. *Journal of Protozoology* 29:136-139.

Sawyer, T.K. (1975a). *Clydonella* n. g. (Amoebidae: Thecamoebidae), proposed to provide an appropriate generic home for Schaeffer's marine species of *Rugipes*, *C. vivax* (Schaeffer, 1926) n. comb. *Transactions of the American Microscopy Society* 94:395-400.

Sawyer, T.K. (1975b). Marine amoebae from surface waters of Chincoteague Bay, Virginia: one new genus and eleven new species within the families Thecamoebidae and Hyalodiscidae. *Transactions of the American Microscopy Society* 94:305-323.

Sawyer, T.K. (1980). Marine amoebae from clean and polluted bottom sediments of the Atlantic Ocean and Gulf of Mexico. *Journal of Protozoology* 27:13-32.

Sawyer, T. & Griffin, J.L. (1975). A proposed new family, Acanthamoebidae n. fam. (order Amoebida) for certain cyst-forming filose amoebae. *Transactions of the American Microscopy Society* 94:93-98.

Schaeffer, A.A. (1926). Taxonomy of the amebas with descriptions of thirty-nine new marine and freshwater species. *Papers from the Department of Marine Biology of the Carnegie Institution of Washington*. Carnegie Institution of Washington, Washington.

Scharf, S., Horn, G.T. & Erlich, H.A. (1986). Direct cloning and sequence analysis of enzymically amplified genomic sequences. *Science* 233:1076-1078.



Schlegel, M. (1991). Protist evolution and phylogeny as discerned from small subunit ribosomal-RNA sequence comparisons. *European Journal of Protistology* 27:207-219.

Schlegel, M. (1994a). Protist evolution and phylogeny as discerned from small subunit ribosomal RNA sequence comparisons. *Journal of Protistology* 27:207-219.

Schlegel, M. (1994b). Molecular phylogeny of eukaryotes. *Trends in Ecology and Evolution* 9:330-335.

Schlegel, M., Elwood, H.J. & Sogin, M.L. (1991). Molecular evolution in hypotrichous Ciliates: sequence of the small subunit ribosomal RNA genes from *Onychodromus quadricornutus* and *Oxytricha granulifera* (Oxytrichidae, Hypotrichida, Ciliophora). *Journal of Molecular Evolution* 32:64-69.

Schmidt, T.M., De Long, E.F. & Pace, N.R. (1991). Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *Journal of Bacteriology* 173:4371-4378.

Schofield, C., Gacesa, P., Price, J.H., Russell, S.J. & Bhoday, R. (1991). Restriction fragment length polymorphisms of enzymically-amplified small-subunit rRNA-coding regions from *Gracilaria* and *Gracilariopsis* (Rhodophyta) - a rapid method for assessing 'species' limits. *Journal of Applied Phycology* 3:329-334.

Scholin, C.A., Anderson, D.M., & Sogin, M.L. (1993). Two distinct small-subunit ribosomal RNA genes in the north american toxic dinoflagellate *Alexandrium fundyense* (Dinophyceae). *Journal of Phycology* 29:209-21.

Schuster, F.L. (1989). Phylum Rhizopoda. In: *Handbook of Protoctista*, ed. Margulis, L., Corliss, J.O., Melkonian, M. & Chapman, D.J., Jones and Bartlett Publishers, Boston.

Seebeck, T. & Braun, R. (1982). Organisation of rDNA in chromatin: *Physarum*. In: *The Cell Nucleus*, ed. Busch, H. & Rothblum, L., Academic Press, New York.

Sieburth, J.McN. (1979). *Sea microbes*. Oxford University Press, New York.

Sieburth, J.McN. (1981). Protozoa bacterivory in pelagic marine waters. In: Heterotrophic Activity in the Sea, ed. Hobbie, J.E. & Williams, P.J.LeB., NATO Conference Series, Plenum Press, New York, London.

Simon, N., LeBot, N., Marie, D., Partensky, F. & Vaulot, D. (1995). Fluorescent *in situ* hybridization with rRNA-targeted oligonucleotide probes to identify small phytoplankton by flow cytometry. Applied and Environmental Microbiology 61:2506-2513.

Singer-Sam, J., Tanguay, R.L. & Riggs, A.D. (1989). Use of Chelex to improve the PCR signal from a small number of cells. Amplifications: a forum for PCR users 3:11.

Singh, B.H. & Das, S.R. (1970). Studies on the pathogenic and non-pathogenic small free-living amoebae and the bearing of nuclear division on the classification of the order Amoebida. Philosophical transactions of the Royal Society, London (B). 259:473-476.

Singh, B.H. & Hanumaiah, V. (1979). Studies on pathogenic and non-pathogenic amoebae and the bearing of nuclear division and locomotive form and behaviour on the classification of the order Amoebida. Monograph No. 1, Association of Microbiologists of India, Indian Journal of Microbiology 1-80.

Sleigh, M.A. (1989). Protozoa and other Protists. Edward Arnold, London.

Smirnov, A.V. & Goodkov, A.V. (1993). *Paradermamoeba valamo* gen.n., sp.n. (Gymnamoebia, Thecamoebidae) - a freshwater amoeba from the bottom sediments. Zoologicheskyy Zhurnal 72:5-11.

Smith, R. McK. & Patterson, D.J. (1986). Analyses of the helizoan interrelationships: an example of the potential trials and limitations of ultrastructural approaches to the study of protistan phylogeny. Proceedings of the Royal Society London B 227:325-326.

Sneath, P.H.A. & Sokal, R.R. (1973). Numerical Taxonomy. W.H. Freeman, San Fransisco.

- Sogin, M.L. (1989). Evolution of eukaryotic micro-organisms and their small subunit ribosomal RNAs. *American Zoologist* 29:487-499.
- Sogin, M.L. (1991). Early evolution and the origin of eukaryotes. *Current Opinions in Genetics and Develoment* 1:457-463.
- Sogin, M.L. & Elwood, H.J. (1986). Primary structure of the *Paramecium tetraurelia* small subunit rRNA coding region: phylogenetic relationships within the Ciliophora. *Journal of Molecular Evolution* 23:53-60.
- Sogin, M.L. & Gunderson, J.H. (1987). Structural diversity of eukaryotic small subunit ribosomal RNAs. *Annals of the New York Academy of Sciences* 503:125-139.
- Sogin, M.L., Elwood, H.J. & Gunderson, J.H. (1986a). Evolutionary diversity of eukaryotic small subunit rRNA genes. *Proceedings of the National Academy of Science USA* 83:1383-1387.
- Sogin, M.L., Miotto, K. & Miller, L. (1986b). Primary structure of the *Neurospora crassa* small unit ribosomal RNA coding region. *Nucleic Acids Research* 14:9540-9540.
- Sogin, M.L., Ingold, A., Karlok, M., Nielson, H. & Engberg, J. (1986c). Phylogenetic evidence for the aquisition of ribosomal RNA introns subsequent to the divergence of the major tetrahymena groups. *EMBO Journal* 5:3625-3630.
- Sogin, M.L., Gunderson, J.H., Elwood, H.J., Alonso, R.A. & Peattie, D.A. (1989). Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. *Science* 243:75-77.
- Spear, B.B. & Prescott, D.M. (1980). Nuclear DNA in normal and refed *Amoeba proteus*. *Experimental Cell Research* 130:387-392.
- Spurr, A.R. (1969). A low-viscosity epoxy resin embedding medium for electron microscopy. *Journal of Ultrastructure Research* 26:31-43.

- Steinbrück, G. Haas, I., Hellmer, K.-H. & Ammermann, D. (1981). Characterisation of macronuclear DNA in five species of ciliates. *Chromosoma* 83:199-208.
- Stumph, W.E., Wu, J.R. & Bonner, J. (1979). Determination of the size of rat ribosomal DNA repeating units by electron microscopy. *Biochemistry* 18:2864-2871.
- Sweeney, R., Chen, L.H. & Yao, M.-C. (1993). Phenotypic effects of targeted mutations in the small-subunit ribosomal-RNA gene of *Tetrahymena thermophila*. *Molecular and Cellular Biology* 13:4814-4825.
- Sweeney, R., Chen, L.H. & Yao, M.-C. (1994). A ribosomal RNA variable region has an evolutionarily conserved essential role despite sequence divergence. *Molecular and Cellular Biology* 14:4203-4215.
- Swofford, D.L. (1981). On the utility of the distance Wagner procedure. In: *Advances in Cladistics*, ed. Funk, V.A. & Brooks, D.R., Proceedings of the First Meeting of the Willi Hennig Society, New York.
- Swofford, D.L. (1993). PAUP: Phylogenetic Analysis Using Parsimony, Version, 3.1. Illinois, Natural History Survey, Champaign, Illinois.
- Swofford, D.L., Olsem, G.J., Waddekk, P.J. & Hillis, D.M. (1996). Phylogenetic inference. In: *Molecular Systematics*, ed. Hillis, D.M., Moritz, C., Mable, B.K., Sinauer Associates Inc., Sunderland, Massachusetts.
- Tan, I.H. & Druehl, L.D. (1993). Phylogeny of the north east pacific brown algal (Phaeophyceae) orders as inferred from 18S rRNA gene sequences. *Hydrobiologia* 261:699-704.
- Tautz, D., Tautz, C., Webb, D. & Dover, G.A. (1987). Evolutionary divergence of promoters and spacers in the rDNA family of four *Drosophila* species: implications for molecular coevolution in multigene families. *Journal of Molecular Biology* 195:525-542.
- Taylor, F.J.R. (1976). Flagellate phylogeny: a study in conflicts. *Journal of Protozoology* 23:28-40.

Tracey, T.E. & Mulcahy, L.I. (1991). A simple method for direct automated sequencing of PCR fragments. *BioTechniques* 11:68-75.

Van de Peer, Y., Hendriks, L., Goris, A., Neefs, J., Vancanneyt, M., Kersters, K., Berny, J., Hennebert, G.L. & De Wachter, R. (1992). Evolution of the basidiomycetous yeasts as deduced from small subunit RNA sequences. *Plant Systematics and Evolution* 170:85-96.

Van de Peer, Y., Neefs, J.-M., De Rijk, P. & De Wachter, R. (1993). Evolution of eukaryotes as deduced from small ribosomal subunit RNA sequences. *Biochemical Systematics and Ecology* 21:43-55.

Van de Peer, Y., Van den Broeck, I., De Rijk, P. & De Wachter, R. (1994). Database on the structure of small ribosomal subunit RNA. *Nucleic Acids Research* 22:3488-3494.

Van de Peer, Y., Nicolai, S., De Rijk, P. & De Wachter, R. (1996). Database on the structure of small ribosomal subunit RNA. *Nucleic Acids Research* 24:86-91.

Vargas, R. & Hattori, T. (1990). The distribution of protozoa amongst soil aggregates. *FEMS Microbiology Ecology* 74:73-78.

Venema, J., Dirksmulder, A., Faber, A.W. & Raué, H.A. (1995). Development and application of an *in vivo* system to study yeast ribosomal RNA biogenesis and function. *Yeast* 11:145-156.

Vila, A., Viril-Farley, J. & Tapprich, W.E. (1994). Pseudoknot in the central domain of small-subunit ribosomal-RNA is essential for translation. *Proceedings of the National Academy of Science USA* 91:11148-11152.

Vilgalys, R. & Gonzalez, D. (1990). Ribosomal DNA restriction fragment length polymorphisms in *Rhizoctonia solani*. *Molecular Plant Pathology* 80:151-158.

Vodkin, M.H., Howe, D.K., Visvesvara, G.S. & McLaughlin, G.L. (1992). Identification of *Acanthamoeba* at the generic and specific levels using the polymerase chain reaction. *Journal of Protozoology* 39:378-385.

- Vogt, V.M. & Braun, R. (1976). Structure of ribosomal DNA in *Physarum polycephalum*. *Journal of Molecular Biology* 106:567-587.
- Vossbrinck, C.R., Maddox, J.V., Friedman, S., Debrunner-Vossbrinck B.A. & Woese, C.R. (1987). Ribosomal RNA sequences suggest microsporidia are extremely ancient eukaryotes. *Nature* 326:411-414.
- Vossbrinck, C.R., Baker, M.D., Didier, E.S., Debrunner-Vossbrinck, B.A. & Shadduck, J.A. (1993). Ribosomal DNA sequences of *Encephalitozoon hellem* and *Encephalitozoon cuniculi*: species identification and phylogenetic construction. *Journal of Eukaryotic Microbiology* 40:354-362.
- Wada, H., Makabe, K.W., Nakauchi, M. & Satoh, N. (1992). Phylogenetic relationships between the solitary and colonial ascidians, as inferred from the sequence of the central region of their respective 18S rRNAs. *Biological Bulletin* 183:448-455.
- Wagner, M., Amann, R., Lemmer, H., Manz, W. & Schleifer, K.H. (1994). Probing activated sludge with fluorescently labelled rRNA targeted oligonucleotides. *Water Science and Technology* 29:15-23.
- Walsh, P.P., Metzger, D.A. & Higuchi, R. (1991). Chelex-100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* 10:506-513.
- Wainright, P.O., Hinkle, G., Sogin, M.L. & Stickel, S.K. (1993). Monophyletic origins of the metazoa: an evolutionary link with fungi. *Science* 260:340-342.
- Ward, D.M., Weller, R. & Bateson, M.M. (1990). 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* 345:63-65.
- Weekers, P.H.H., Gast, R.J., Fuerst, P.A. & Byers, T.J. (1994). Sequence variation in small-subunit Ribosomal RNAs of *Hartmannella vermiformis* and their phylogenetic implications. *Molecular Biology and Evolution* 11:684-690.

- Welsh, J. & McClelland, M. (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* 18:7213-7218.
- Whittaker, R.H. (1959). On the broad classification of organisms. *Quarterly Review of Biology* 34:210-226.
- Wiaderkiewicz, R. & Ruiz-Carrillo, A. (1987). Mismatch and blunt to protruding-end joining by DNA ligases. *Nucleic Acids Research* 15:7831-7848.
- Wilcox, L.W., Fuerst, P.A. & Floyd, G.L. (1993). Phylogenetic relationships of four charophycean green algae inferred from complete nuclear-encoded small subunit rRNA sequences. *American Journal of Botany* 80:1028-1033.
- Wild, M.A. & Sommer, R. (1980). Sequence of a ribosomal RNA gene intron from *Tetrahymena*. *Nature* 283:693-694.
- Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A. & Tingey, S.V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* 18:6531-6535.
- Winnepenninckx, B., Backeljau, T. & De Wachter, R. (1994). Small ribosomal RNA and the phylogeny of Mollusca. *The Nautilus* 2:98-110.
- Woese, C.R. (1987). Bacterial evolution. *Microbiological Reviews* 51:221-271.
- Woese, C.R. & Gutell, R. (1989). Evidence for several higher order elements in ribosomal RNA. *Proceedings of the National Academy of Science USA* 86:3119-3122.
- Woese, C.R., Magrum, L.J. & Fox, G.E. (1978). Archaeobacteria. *Journal of Molecular Evolution* 11:245-252.
- Woese, C.R., Gutell, R., Gupta, R. & Noller, H.F. (1983). Detailed analysis of the higher order structure of 16S-like ribosomal ribonucleic acids. *Microbiological Reviews* 47:621-669.

Woese, C.R., Kandler, O. & Wheelis, M.L. (1990). Towards a natural system for organisms: proposals for the domains Archea, Bacteria and Eucarya. *Proceedings of the National Academy of Science USA* 87:4576-4579.

Wohlfarth-Botterman, K.E. (1960). Protistenstudien X. Licht- und elektronenmikroskopische untersuchungen an der amöbe *Hyalodiscus simplex* n. sp. *Protoplasma* 52:58-107.

Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G.J. & Woese, C.R. (1985). Mitochondrial origins. *Proceedings of the National Academy of Science USA* 82:4443-4447.

Yang, Z. (1993). Maximum likelihood estimation of phylogeny from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution* 39:306-324.

Yasunobu, K.T. & Tanaka, M. (1975). The evolution of iron-sulphur protein containing organisms. *Systematic Zoology* 22:570-589.

Yonath, A., Bennett, W., Weinstein, S. & Wittmann, H.G. (1990). Crystallography and image reconstructions of ribosomes. In: *The Ribosome: Structure, Function and Evolution*, ed. Hill, W.E., Dahlberg, A., Garrett, R.A., Moore, P.B., Schlessinger, D. & Warner, J.R., American Society for Microbiology, Washington D.C.

Zachau, H.G., Dutting, D., Feldmann, H., Melchers, F. & Karau, W. (1966). Serine specific transfer ribonucleic acids. XIV. Comparisons of nucleotide sequences and secondary structure models. *Cold Spring Harbor Symposium on Quantitative Biology* 31:417-424.

Zarda, B., Amann, R., Wallner, G. & Schleifer, K.-H. (1991). Identification of single bacterial cells using digoxigenin-labelled rRNA-targeted oligonucleotides. *Journal of General Microbiology* 137:2823-2830.

Zhu, X., Wittner, M., Tanowitz, H.B., Cali, A. & Weiss, L.M. (1994). Ribosomal RNA sequences of *Enterocytozoon bieneusi*, *Septata intestinalis* and *Ameson michaelis*: phylogenetic construction and structural correspondence. *Journal of Eukaryotic Microbiology* 41:204-209.



Zuckerkandl, E. & Pauling, L. (1965). Molecules as documents of evolutionary history. *Journal of Theoretical Biology* 8:357-366.

Zuker, M. (1989). Computer prediction of RNA structure. *Methods in Enzymology* 180:262-288.

## Appendix I: Culture media

Recipes for liquid and agar media used for the cultivation of amoebae and bacteria are described below. All media were sterilized by autoclaving for 20 mins at 15 lb per square inch pressure.

### **Amoeba saline solution (AS)**

AS Stock A	5 mls
AS Stock B	5 mls
Distilled water to	1 litre

AS Stock A		AS Stock B	
NaCl	12.0 g	Na <sub>2</sub> HPO <sub>4</sub>	14.2 g
MgSO <sub>4</sub> .7H <sub>2</sub> O	0.4 g	KH <sub>2</sub> PO <sub>4</sub>	13.6 g
CaCl <sub>2</sub> .6H <sub>2</sub> O	0.6 g	Distilled water to	500 mls
Distilled water to	500 mls		

### **Malt yeast 75% seawater agar (MY75S)**

Malt extract	0.1 g
Yeast extract	0.1 g
Agar technical No.3 (Oxoid)	12 g
Filtered seawater	750 mls
Distilled water	250 mls

### **Non-Nutrient Agar (NNA)**

Agar technical No.3 (Oxoid)	12 g
Amoeba saline	1 litre

### **Nutrient Agar (NA)**

Nutrient agar (Oxoid)	28 g
Distilled water to	1 litre

### **PPG Liquid Media**

Protease peptone	15 g
Glucose	18 g
Amoeba saline solution	1 litre

### **SOC Liquid media**

Tryptone	20 g
Yeast extract	5 g
NaCl	0.5 g

The solutes were dissolved in 950 mls of distilled water, 10 mls of 250 mM KCl was added, the solution was adjusted to pH 7.0 with 5 M NaOH, made up to a volume of 1 litre with distilled water and sterilized by autoclaving. The media was allowed to cool and 5 ml of sterile 2 M  $\text{MgCl}_2$  and 20 ml of filter sterilized 1 M glucose were added prior to use.

